

France sleepwalks into chaos

The success of the extreme right in the first round of the French presidential elections serves as a warning to all responsible citizens — scientists included — not to disengage from the political process.

A cartoon on the front of *Le Monde* said it all: an airliner with the face of extreme right-winger Jean-Marie Le Pen crashing into twin towers labelled Jacques Chirac, the centre-right French president, and Lionel Jospin, the socialist prime minister. That the latter two men would defeat the 14 other candidates in the first round of France's presidential elections was regarded as a foregone conclusion. The news that Le Pen will run head-to-head against Chirac in the second round has rocked French society to the core.

Le Pen's strong showing does not mean that France is seriously flirting with government by the extreme right: the combined vote of the mainstream right and left will keep Le Pen from office, and hand victory to Chirac. Rather, the first-round results are the manifestation of a widening gap between the French people and its ruling political class. Turned off by a campaign in which neither mainstream candidate engaged in a meaningful debate on such issues as France's future in Europe, many citizens registered a protest vote for a minority candidate, or simply stayed at home.

The shock of finding Le Pen's name on the ballot paper for the second round presents both a danger and an opportunity. There is a risk that protests against Le Pen will spill over into violence, playing into the extremists' hands. More optimistically, the shock wave could shake France out of its slumber, and displace the uninspiring faces that have dominated the French political scene for three decades. The humiliated Jospin has already said that he will quit politics.

Needless to say, policies on science and innovation were largely absent from the electoral debate. Yet they are key to the economic and intellectual future of any modern country — and France's staid, bureaucratic and inefficient research system is in urgent need of

attention. Today's scientific enterprise needs a flexible, highly mobile workforce. The French system, in which most scientists are civil servants who can spend their entire careers attached to one research unit, is ill-equipped to provide this. French science needs a postdoc system to encourage mobility between research groups; universities need a shake-up to encourage more productive interactions with the public research laboratories that they host, plus an injection of funds to free scientists from long teaching hours; and research labs need to be lifted from the treacle-like bureaucracy of French public administration.

Once Le Pen has been seen off, French researchers have everything to gain from re-engaging with the political process and putting science back on the agenda. Responsibility for science policy lies not with the president, but with the prime minister's government, and the parliamentary elections that will determine who forms that government will take place in June.

Recent governments of left and right have not served French science well. In the mid-1990s, the right slashed budgets. And while the socialists restored funding, they have let the underlying problems of French science rumble on. Jospin's first science minister, Claude Allègre, realized the need for reform, but his undiplomatic style and poorly thought-through plans delivered little beyond a series of rows with the organizations he was seeking to change. His successor, Roger-Gérard Schwartzberg, has avoided such conflicts. But that is about the sum of his achievement.

France deserves better. The country's scientists should now play their part in putting in place a government that is up to the task of leading one of Europe's great nations in the twenty-first century. ■

Maintaining the climate consensus

The election of a new chair for the Intergovernmental Panel on Climate Change has left wounds that the victor must heal.

The circumstances under which the Indian energy economist Rajendra Pachauri won the chair of the Intergovernmental Panel on Climate Change (IPCC) are not auspicious. As soon as he was elected, Pachauri was denounced in *The New York Times* by former US vice-president Al Gore as "the 'let's drag our feet' candidate". For the new head of the body that is supposed to advise the world's governments on the complexities of global warming, things can only get better.

Previously, consensus had emerged on the IPCC's leadership. But an orchestrated campaign by the US administration and the fossil-fuel lobby forced the vote on 19 April in which Pachauri defeated the incumbent, atmospheric scientist Robert Watson, by 76 votes to 49.

Climate researchers appreciated the way in which Watson defended their findings from politically motivated attacks during his tenure. Many will now be wary of Pachauri, who appears to have tarnished his reputation by collaborating with those whose objective was to ditch Watson.

But Pachauri has the credentials to make a go of his main role, which is to build confidence in the impartiality of the IPCC's advice.

His expertise in energy policy and economic development is central to the panel's mission, and his involvement with the oil industry — he is a director of the Indian Oil Corporation — may help to establish more credibility with business interests. And after two leaders from rich nations (Watson's predecessor, meteorologist Bert Bolin, was Swedish) it is good for the IPCC to be run by someone from a developing country.

Pachauri will have to establish his credibility with scientists quickly, however. Researchers give their time to the panel for free, and they need to feel that the chair will back them up. Pachauri must reassure the IPCC's rank and file that he will fend off any attempts to bully the panel into watering down its findings.

He should also be thinking about negotiations on greenhouse-gas emission targets under the Kyoto Protocol from 2013. Talks begin in 2005, two years before the IPCC's next assessment is due. An interim report from the IPCC would inform the negotiations about the latest science. Such a document can only be requested by the panel's member states, but Pachauri should make it clear that his teams are ready and willing to provide it. ■





Shine on
Upgrade gives Vulcan
the world's most
intense laser pulses
p775



Seeing stars?
Recovered disc opens
debate about Bronze
Age astronomy
p776



Bad reaction
UK students avoid
studying chemistry
at university
p777



Change of heart
South Africa drops
opposition to
anti-HIV drug
p778

Allegations of rushed proposals mar disease fund's first awards

Erika Check, Washington

Members of the new Global Fund to Fight AIDS, Tuberculosis and Malaria were set to meet in New York this week to decide how best to distribute its first round of funding.

But its governing board is already facing some tough questions about how it has managed the ambitious international project during its debut year.

The fund was launched last July, when leaders of world's eight most powerful nations pledged to contribute US\$1.3 billion between them for public-health projects — including some applied research — aimed at countering the three diseases, which between them kill about six million people each year. Further public and private donations have since built the fund up to \$1.9 billion.

A technical review panel vets grant proposals before the governing board makes the final decisions. Fund supporters say these provisions should make for a rational, fair and expeditious grant-application process.

But some scientists and government officials charge that the fund was slow to issue clear guidelines for grant proposals. Interested parties were then given only two months to develop proposals for the first round of grants, worth up to \$400 million,



Rapid response: a global fund to fight diseases such as malaria is struggling to meet its own schedules.

to be disbursed at the New York meeting.

"We probably did it too fast and didn't give enough information out," concedes Bill Steiger, a health-department official involved in running the fund in the United States.

The fund has also not determined how it will evaluate the success of projects that it

does support, despite requests from anxious donor governments that it does so.

Scientists say that the formation of the technical review panel — set up just weeks before it met in Geneva in March to look at 322 grant proposals — was rushed.

"People are worried that if they rush the

Animal video nasty sets fur flying over exemption bill

Rex Dalton, San Diego

An animal-rights group has infiltrated a leading US research university and unearthed what it claims is a mountain of incriminating video evidence of animal mistreatment.

People for the Ethical Treatment of Animals (PETA) hopes to use the evidence to overturn a current congressional effort to exempt rodents and birds from tighter restrictions on research.

An employee spy with a hidden camera videotaped neuroscience experiments carried out at the University of North Carolina at Chapel Hill from last December until March. The videotape shows possible violations of accepted practices for

procedures using mice and rats, together with researchers' on-camera explanations of why they were breaking the rules.

The video shows neurological procedures on rodents that may not have been properly anaesthetized, experiments on rodents with diseases and injuries, and rodents that were still alive after being discarded as waste.

Senator Jesse Helms (Republican, North Carolina) has introduced an amendment to an agricultural bill that would block the extension of the Animal Welfare Act to cover rodents and birds. The act includes stringent standards for research on higher animals such as dogs, cats and primates.

The Helms amendment has been supported by many research organizations, who fear that the extension could block scientific projects and raise costs.

Tony Waldrop, a neurophysiologist and vice chancellor for research at the University of North Carolina, says that the university is conducting an investigation into the events shown on the videotape. "How the allegations came forward is not important," says Waldrop. "If things aren't done correctly, we will deal with it." Research on rodents and birds follow guidelines set by the National Institutes of Health, which says that it has not yet begun its own inquiry into the video. ■

♦ www.peta-online.org

money into the hands of countries that are not prepared to spend it effectively, it's going to be just as bad as not spending it at all," says Richard Chaisson, director of the Center for Tuberculosis Research at Johns Hopkins University in Baltimore. Chaisson was invited to serve on the review panel but says that he was unable to do so because of the short notice he was given.

"We have to get it right this time," says Gail Cassell, vice-president for infectious diseases at Eli Lilly, an Indiana-based drug company. "I fear that, because we are expected to do it so rapidly, we won't."

The fund's supporters say they had to move quickly to meet the expectations of donors. "There has been enormous pressure for us to move fast and to show that this can work," says Anders Nordstrom, the fund's interim director. He says that most of the criticism comes from scientists and officials in rich, donor countries, not from the poor countries that have to deal with the diseases.

And while the fund is being attacked for imperfections resulting from its haste to swing into action, other critics argue that its war chest is too puny to tackle the diseases. They point out that Kofi Annan, the secretary general of the United Nations, who pushed for the creation of the fund, has said that between \$7 billion and \$10 billion in extra aid is needed to combat AIDS alone.

"The amount of money that donors have contributed so far is quite unimpressive," says Amir Attaran of Harvard University's John F. Kennedy School of Government. But government officials say that the private and public sectors will increase their investment once they see that their initial money has been wisely spent. ■

Climate panel unsettled by public battle for top job

Jim Giles, London

The Intergovernmental Panel on Climate Change (IPCC) elected a new chair last week — but the nature of the vote has raised questions over the panel's independence.



IPCC's new chair: **Rajendra Pachauri.**

Rajendra Pachauri, director of the Tata Energy Research Institute in New Delhi and vice-chair of the IPCC, beat current chair Robert Watson by 76 votes to 49 in the election held in Geneva on 19 April.

Pachauri, an engineer and economist by training, is a respected expert in the economics of development.

But the public nature of the contest — which is usually settled by consensus during behind-the-scenes negotiations — has posed some uncomfortable questions for the panel, set up in 1988 to advise the world's governments on climate-change issues.



Robert Watson: **expected by many to win.**

Watson, an ecologist who is chief scientist at the World Bank, had been expected by many climate researchers to serve a second, six-year term. But he was not renominated for the position by the United States, which backed his initial candidacy in 1996 (see *Nature* 416, 251; 2002).

He eventually received his nomination from Portugal and New Zealand, after the official 15 March nomination deadline had passed.

Watson's former role as associate director for environment at the White House Office of Science and Technology Policy under President Bill Clinton is thought to have deterred the Bush administration from putting him forward. But evidence has also emerged that ExxonMobil, the US oil company, lobbied the administration not to renominate Watson.

The campaign against him continued at the election meeting in Geneva. "Oil-industry representatives were there lobbying for Pachauri," says Bert Metz, a climate-policy expert at the National Institute of Public Health and the Environment in Bilthoven, the Netherlands, and co-chair of the IPCC working group on mitigating climate change.

Despite the split, IPCC scientists are keen to stress that Pachauri will have their goodwill. Michael Grubb, a climate-change and energy-policy specialist at Imperial College, London, rates Pachauri as extremely capable. "The only worrying thing is the manner in which he was elected," he says.

Pachauri will now have to guide the IPCC towards its fourth climate-change report, due in 2007. Researchers say that the panel's three working groups, which cover climate science, the impact of climate change and the steps that can be taken to mitigate it, respectively, are all in good shape. "The election won't cause lasting damage," says John Houghton of the Hadley Centre for Climate Prediction and Research in Bracknell, west of London, who stepped down as chair of the climate-science group at the elections. "All the groups are in very good hands." ■

Japan plots course for university on naval-base island

David Cyranoski, Tokyo

Japan is pulling into focus its plans for a graduate university that will conduct all of its research and lectures in English.

The first meeting of an international review committee for the project is set to take place on 26–27 April in Santa Monica, California, where leading US scholars, including David Baltimore, president of the California Institute of Technology, will offer their advice.

The university is to be based on the poor, southern island of Okinawa, and will focus on biosciences and information technology. Supporters of the project hope that it will energize a local economy that has largely depended on a locally unpopular US naval base. But some researchers suggest that the

island's remote location will make the facility less attractive to potential recruits.

The plan is being championed by Koji Omi, an influential politician who is minister for both the Okinawa region and for national science and technology policy. He says that a lack of English language skills is holding back Japan's science. "I couldn't believe it, but you can get a PhD in a science from many of Japan's universities without knowing English," he says.

Omi wants the college to be a world-class facility. "It will be a university at the very highest level internationally," he says. He hopes to attract 500 research staff and 500 graduate students — more than half of them from foreign countries.

But some researchers are sceptical. "I

have serious doubts about whether people will come," says Shigeru Ohde, a marine chemist at the University of the Ryukyus outside Okinawa's capital, Naha. "Why would they come here to study life sciences or IT? Those fields can be studied anywhere." Critics also fear that the project will run out of steam should Omi leave office.

The plan is expected to feature in next year's budget, which will be presented by the government in December. Current estimates for construction costs run from ¥20 billion to ¥80 billion (US\$150 million to \$600 million), plus ¥20 billion annually for operations. The university is likely to be established as a private institution with strong government support, avoiding the red tape involved in setting up a new public university. ■

Astronomers get physical for future plans

Geoff Brumfiel, Albuquerque

A satellite to study the radiation left over by the Big Bang and a laboratory deep underground are among the projects endorsed by a US survey of the interface between astronomy and physics.

Researchers working at this interface are hoping that the survey, released at a joint meeting between the American Physical Society and the American Astronomical Society on 21 April, will help to build support for their work.

The recommendations follow a two-year study undertaken at the request of Dan Goldin, then the director of NASA, after he identified astrophysics as a key area for collaboration between NASA and other agencies (see *Nature* **399**, 509; 1999). The report also endorses Goldin's vision of stronger links between NASA and Earth-bound science agencies, to help get some of the projects off the ground.

Each of three instruments newly endorsed by the National Academy of Sciences' report is designed to probe a different question in cosmology. One of them, a 2-metre diameter wide-field space telescope, would probe the 'dark energy' that seems to be playing a role in the expansion of the Universe. A second satellite would look for polarization in the cosmic microwave background left over from the Big Bang. Researchers believe that this might help to explain how gravity waves influenced the rapid inflation of the early Universe.

Experiments conducted in the underground laboratory would include a search for the decay of protons — an extremely rare event that high-energy physicists have predicted but never observed — as well as probing the neutrino, a hard to detect particle whose recently discovered mass

poses a challenge to the standard model of physics.

This year, construction of the underground laboratory has become an issue in a tight Senate election race in South Dakota, home of the disused Homestake goldmine where the facility would probably be built (see *Nature* **415**, 105; 2002).

The report also adds its support to three projects already endorsed in the National Academy of Sciences' 2001 decadal review of astronomy: the Laser Interferometer Space Antenna, which would detect gravity waves; Constellation-X, a group of four satellites to probe the edges of black holes; and the Large-aperture Synoptic Survey Telescope, which would search for so-far undetected dark matter in deep space.

All of these programmes, the authors suggest, should be run through a new inter-agency initiative between NASA, the Department of Energy and the National Science Foundation. The report does not say how much such an initiative would cost, but Michael Turner, an astrophysicist at the University of Chicago, who chaired the panel that wrote the report, says that it might spend "something like a billion dollars" over 10 years.

Michael Lubell, director of public affairs at the American Physical Society, says he is optimistic that the report will help to raise awareness of the physical sciences both among the general public and in the US Congress, which will ultimately decide which of the projects will proceed. ■

Laser gears up for star role

Natasha McDowell, London

Astrophysicists will soon be able to deploy a range of lab experiments to test the predictions of some of their theories.

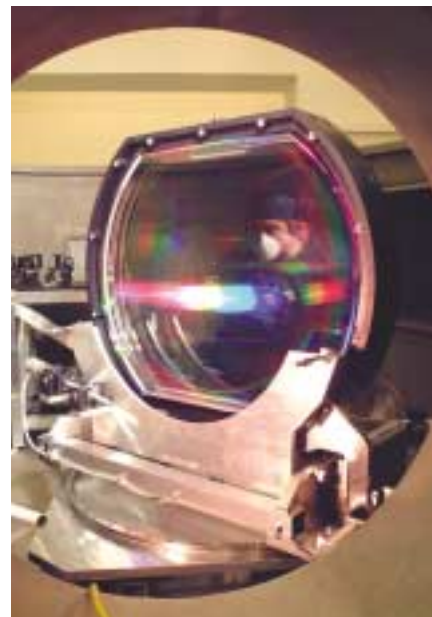
The upgraded Vulcan laser, which was inaugurated this month at the Rutherford Appleton Laboratory near Oxford, will offer researchers the most intense laser pulses in the world. When trained on a target, these pulses will create temperatures and pressures comparable to those found inside stars.

"Vulcan will provide us with the experimental data to check the theoretical calculations that have come from observations," says Steven Rose, an astrophysicist at the University of Oxford.

Vulcan's upgrade will see it emit near-infrared pulses containing a relatively modest 500 joules of energy. But their ultrashort life — the pulses last just 500 femtoseconds (5×10^{-15} seconds) — means that they offer a peak power of one petawatt (10^{15} watts). This exceeds the power produced by typical high-energy machines such as the OMEGA laser at the University of Rochester, New York, which emits longer pulses containing 40,000 joules.

"The increased intensity of Vulcan will drive research into new territories," says Chris Edwards, manager of the Vulcan upgrade. "It is sure to produce new surprises." Initial tests at the facility are scheduled to begin next month, and it should open to users in November.

Vulcan's beam will be focused onto a spot just 10 micrometres across, producing what its operators say will be one of the most intense laser beams ever generated in a laboratory. It will be used for a range of

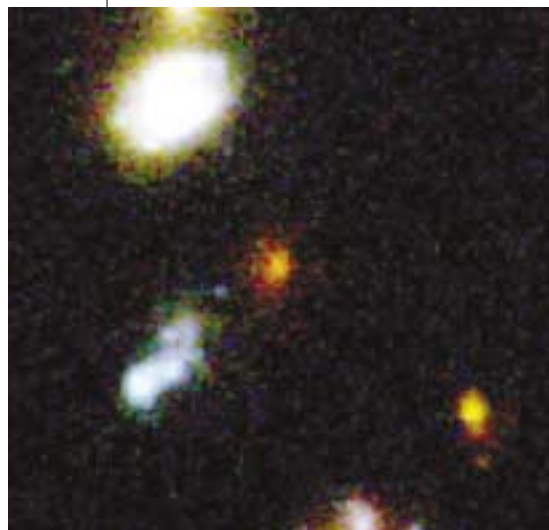


Getting a grip: the Vulcan laser will help astrophysicists to test their theories.

experiments — Rose, for example, plans to use it to study iron, a constituent of the Sun that plays an important role in conducting energy outwards from the fusion reactions at the star's core. It will also be involved in research into a 'fast ignition' approach to nuclear fusion energy, which uses the intense pulse to initiate fusion reactions.

Vulcan will be used primarily by British and European researchers, but access has also been arranged for scientists in the United States. Similar high-intensity lasers are under development in Japan and Germany, and researchers are seeking funding for a facility in France. ■

A. RIESS ET AL./STC/NASA



Observations of this supernova by the Hubble Space Telescope hint at the role of dark energy.

Melting ice triggers Himalayan flood warning

Natasha McDowell, London

Melting glaciers high in the Himalayan mountains are creating unstable lakes that threaten to burst their banks, endangering the lives of tens of thousands of people, according to a United Nations report.

Although not new, the frequency of such floods has risen in the past 30 years. The increase is blamed on global warming by the report, which was written by researchers at the United Nations Environment Programme (UNEP) and the International Centre for Integrated Mountain Development, based in Kathmandu, Nepal.

Ground surveys and satellite images have identified 44 lakes in Nepal and Bhutan as being in danger of bursting the natural dams constraining them within 5–10 years. Data from 49 monitoring stations across Nepal indicate that air temperatures in the area are on average 1 °C higher than they were in the 1970s.

In Bhutan, some melting glaciers are retreating by 30–40 metres a year, according to the report, and in Nepal's Dolakha District the glacier feeding the Tsho Rolpa lake has retreated by up to 100 metres a year. The lake has grown sixfold since the 1950s, putting 10,000 people and a hydroelectric power station at risk.

Glacial lakes form behind moraines — dams of rocks, debris and ice that are



Water works: operations under way at Tsho Rolpa lake, which has grown sixfold since the 1950s.

left behind as the glacier retreats. These can quickly collapse; a small leakage as the ice melts is easily enlarged, and escaping water effectively eats the moraine from the inside out.

"Water was lapping the top of Tsho Rolpa's moraine and we didn't know how much time we had before it would burst," says John Reynolds, a geophysical engineer at Reynolds Geo-Sciences, a consultancy based in Wales, who is advising the Nepalese government on Tsho Rolpa.

Initial piping to siphon off the water proved insufficient to deal with the rising levels at Tsho Rolpa, Reynolds says. So two years ago, an emergency drainage channel was cut through the moraine to lower the

water level by 3.5 metres. This was a tough job, he points out — the lake lies at an altitude of 4,500 metres and is 60 kilometres from the nearest road. Materials and equipment had to be brought in by porters or helicopters.

The work, which was paid for by the Dutch government, reduced the lake's volume by one-sixth and has allowed part of the moraine to dry out, thereby stabilizing it. But now two large ice blocks in the moraine are melting and the water level needs to be lowered further.

Tsho Rolpa is just one of many lakes facing the same problem. Klaus Toepfer, UNEP's executive director, described the problem as "another compelling reason to act to reduce emissions of carbon dioxide and other greenhouse gases".

REYNOLDS GEO-SCIENCES LTD

Twice-rescued Bronze Age disc goes on public view

Marion Kerstholt, Munich

A bronze disc depicting the Sun, Moon and stars, which went on public exhibition in Germany this month, has sparked a debate about the astronomical expertise of

Bronze Age Europeans. Intrigued archaeologists and astronomers can thank the police, who recovered the artefact after it was stolen from a site at Sangerhausen in the state of Saxony Anhalt, 100 kilometres west of Leipzig.

The gold-decorated disc is on show at the State Museum of Prehistory in Halle. Some 3,600 years old, it may be the first 'scientific' representation of the sky of central Europe, claims Wolfram Schlosser, an astronomer at Ruhr University in Bochum. He believes that the disc represents the constellation Pleiades.

At the time the disc was made, Pleiades was not visible from central Europe in spring. So the constellation's disappearance might have been used to herald the season's arrival. Alix Henseler, an expert on the Bronze Age at the Museum of Prehistory in Berlin, says the disc suggests that "people living in middle Europe in the Bronze Age were able to draw seasonal conclusions from observations of stars".

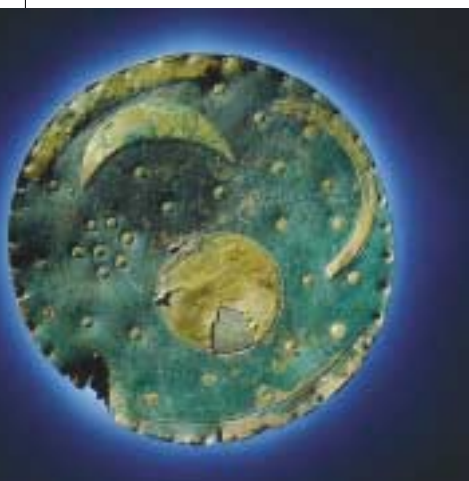
But other astronomers are not convinced. "The bronze disc is clearly a remarkable object, but as an astronomical piece it is very

primitive," says Owen Gingerich, an expert in the history of astronomy at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. He disputes whether the pattern of stars represents a specific constellation.

Experts can count themselves lucky that they are able to debate the object's significance at all. In 1998, two amateur archaeologists excavated the disc and sold it to a dealer for 15,000 euros (US\$13,300) — although under German law all such finds belong to the federal republic. The dealer unsuccessfully tried to sell it to the Berlin museum.

A go-between, who later claimed that she had wanted the disc to stay in Germany, then sold it to a Düsseldorf schoolteacher. Believing that Harald Meller, chief of Saxony Anhalt's department of archaeology, might want to buy the disc for the state, the go-between invited him to meet the teacher in February this year. But Meller was already working closely with German police, and both the teacher and the go-between were arrested. They have yet to be charged.

H. J. LIPTAK/SAXONY ANHALT DEPT ARCHAEOLOG



Astronomical claim: did Bronze Age people identify the seasons by the stars?

Poor nations' crop research hurt by Japanese cutbacks

David Cyranoski, Tokyo

With recent breakthroughs such as the draft rice genome sequence, agriculture researchers should have plenty to look forward to. But the research centres that are best placed to tap that promise are facing a funding crisis this spring.

Japan, a major supporter of such research, is planning to cut funding for the Consultative Group on International Agricultural Research (CGIAR) by almost half, from last year's ¥3.6 billion (US\$29 million). The CGIAR's 16 institutes form the main global network of research centres addressing the agricultural needs of poor countries.

The impact will be felt immediately at the CGIAR's prestigious International Rice Research Institute (IRRI) near Manila in the Philippines. "Japan had been our largest and most faithful donor," laments Ron Cantrell, the institute's director.

The impending cut led IRRI, at a board meeting earlier this month, to reduce staff levels by over 20% to 785. In addition, some projects will be scaled back. The cuts will not affect the maintenance of the world's largest collection of rice germ plasm, which is kept at the institute, Cantrell says.

Other members of the CGIAR are also feeling the pinch. The International Potato Center (CIP) in Lima, Peru, expects to lose about US\$700,000 in Japanese funding and will lay off one-fifth of its staff. Research on late blight and the breeding of new pest-resistant varieties will suffer, says Hugo Li Pun, a deputy director at the CIP. Li Pun has also heard rumours that European nations are set to cut donations.

The situation is especially frustrating because so much data is now available for agricultural research, says Cantrell. "The world is bursting with new technologies," he says. "We are really on the cusp of something exciting."



Chemistry caught in crisis catalysed by student apathy

David Adam, London

Researchers, politicians and teachers are warning of an impending crisis in British chemistry after a government report highlighted a dramatic collapse in the number of university students pursuing the discipline.

Chemistry — like the rest of the physical sciences — is currently struggling to attract degree students all over the world, according to professional societies and others. But the shortfall and the reasons for it in the United Kingdom were brought into sharp relief on 15 April, when the *SET for Success* report was delivered to the British government. It was written by a panel chaired by Oxford University physicist Gareth Roberts.

The Royal Society of Chemistry, the British professional society for chemists, says that the survival of at least a dozen university chemistry departments is in doubt because of a lack of government funding for their research programmes.

"It's very worrying and some action has to be taken," says Brian Iddon, formerly an organic-chemistry researcher at Salford University, and now Labour member of parliament for Bolton South East and a member of the Science and Technology Select Committee. "The numbers entering chemistry have simply collapsed."

David Giachardi, chief executive of the Royal Society of Chemistry, says that recruitment difficulties are now "a worldwide problem". He says that the issue is causing concern in the United States, and the American Chemical Society confirms that the number of chemistry graduates there is in decline.

In Britain, the reasons for the trend are not hard to find. "All of my A-level chemistry class are going to university but none want to do chemistry," says teacher Zoe Thorn at Saffron Waldon County High School in Essex. More students are using chemistry as a route to vocational degrees such as medicine, she says — a situation that could worsen if the limits on medical student numbers in the United Kingdom are raised. Recent advances in molecular biology and genomics are also drawing students to life sciences and away from chemistry.

Chemistry is not the only subject in which alarm bells are ringing. Applications to study all the physical sciences have taken a nosedive in Britain and elsewhere in recent years. The total number of students entering degree courses rose by 12% between 1995 and 2000, according to the Roberts report. But those studying physics and engineering dropped by 7% and the numbers starting chemistry degrees fell by 16%.

The situation could get worse before it



Lack of attraction: students are becoming less inclined to study chemistry at university.

gets better. Officials at the Royal Society of Chemistry privately warned the government earlier this year that the decision by the Higher Education Funding Council for England to cut funding for mid-ranking institutions (see *Nature* **415**, 465; 2002) leaves up to a dozen chemistry departments in England vulnerable to closure.

Few observers would dispute that supply and demand should play a role in restructuring Britain's university system. But the society warns that students who choose to live at home and attend their local universities, because of the rising costs of attending, could soon find it impossible to study chemistry in certain regions, particularly in southern England.

Many heads of mid-ranking British chemistry departments contacted by *Nature* were concerned about falling applications and nervous about the future of their departments, but few would voice their fears openly. One department about to be subsumed by biology and sack several researchers is, its chair says, "building on its interdisciplinary strengths". Another department head, whose physical-chemistry teaching laboratory is being handed over to sports science, admits merely that it is "redistributing resources".

Chemists remain hopeful that the current retreat of students from the discipline may prove to be cyclical, rather than permanent. In Germany in the mid-1990s, for example, the number of chemistry students collapsed by some 60% — but according to Kurt Begitt, director for education and employment with the German Chemical Society, numbers have recovered after an education effort and evidence of improved job opportunities for chemistry graduates.

♦ www.hm-treasury.gov.uk

S. & R. GREENHILL



Drug demand: campaigners have called for South Africa to grant women access to HIV therapies.

N. NKOZI/AP

South African leader relents in storm over HIV drugs for women

Cape Town South Africa's president, Thabo Mbeki, seems to have backed down from his controversial stand against the use of drugs to combat HIV.

Mbeki had previously allowed the drug nevirapine to be made available to only a small number of pregnant women with the virus. Nevirapine reduces the risk of mother-to-child transmission of HIV, but Mbeki had argued that its side-effects outweigh its benefits — a claim that is denied by experts. Pressure has mounted on the government over the past 12 months. Public-health groups launched legal challenges to the government last August, and Mbeki's predecessor, Nelson Mandela, has publicly criticized the policy.

Health minister Manto Tshabalala-Msimang announced last week that nevirapine will be made available to pregnant women in all state hospitals by the end of the year. The therapy will also be provided to rape victims with immediate effect.

University misconduct probe backs under-fire researcher

Sydney The University of New South Wales has found no evidence to support allegations that one of its researchers committed scientific fraud and mismanaged funds.

The university found itself in the spotlight earlier this month when one of its researchers — immunologist Bruce Hall, head of the Division of Medicine at Liverpool Hospital in Sydney — was accused of misappropriating research funds, fabricating data, bullying colleagues and wrongfully including or deleting the names of authors on papers. The allegations, made by three members of Hall's lab, had been under investigation by the university since last year, but came to national attention after the complainants' case was broadcast on an Australian Broadcasting Corporation radio programme.

The university's report, released on

18 April, found no evidence of fraud or mismanagement of funds, but did note errors in scientific reporting and attribution of authorship, as well as inadequacies in communication and workplace behaviour. The report also identified several unresolved issues relating to scientific misconduct and fraud which will now be examined by an independent external enquiry.

Celera leaves genomic jewel to sister outfit

Washington Celera Genomics, the company behind the privately funded human genome sequence, this week further distanced itself from its original mission of being a commercial provider of genetic information. Applera, Celera's parent company, is shifting the sales and marketing of the genomics database to another of its companies — Applied Biosystems of Foster City, California.

Celera Genomics is now focusing on drug discovery — one of the reasons for the recent departure of its president, Craig Venter. Kathy Ordoñez, currently president of Celera Diagnostics, will take over from him. Ordoñez joined Celera in 2000 after several years in the pharmaceutical industry.

Celera Genomics' database has recently become profitable, but industry experts have speculated about the long-term viability of a commercial genomics database, as public databases provide similar information. DoubleTwist, a Californian company that provided tools to view annotated version of the genome, went out of business last month (see *Nature* **416**, 357; 2002).

Protein predictor bags role as political point of contact

Sydney Thomas Barlow, a molecular biologist, has been appointed as science adviser to the Australian science minister, Brendan Nelson.

Barlow studied protein-structure prediction and small-molecule drug design at the University of Oxford before moving to the Massachusetts Institute of Technology, where he developed microarray data visualization tools. He will now be the federal government's point of contact for science bodies and agencies across the country, and, Nelson says, will "ensure that the voice of science is heard loud and clearly in my office".

Barlow's top priorities will be the review of the higher-education sector and methods for prioritizing research.

Japan's scientists fear revolution in schools

Tokyo Moves by Japanese education officials to ease the pressure on the country's children have forced evolution off the national school curriculum.

The subject will no longer be included

in the middle-school curriculum for 12–15-year-olds, and will be optional for 15–18-year-olds in high schools. Maths and other sciences are also being scaled back, as the total number of hours spent in classes is cut. Officials say that reducing the workload will increase the creativity of students, and will stimulate interest in increasingly unpopular science subjects.

But scientists fear that the ministry is underestimating evolution's significance. "Without evolutionary theory, the effort to sequence the genome will only give you a catalogue," wrote Mariko Hasegawa, an evolutionary biologist at Waseda University in Tokyo, in an editorial in the *Asahi Shimbun* newspaper last week. The education ministry has denied rumours — which have been seized upon by private schools — that the new, relaxed mathematics curriculum will approximate π as 3 rather than 3.14.

Smallpox precautions spark 'cash for contract' storm

London British politicians are probing the government's decision to award a £32-million (US\$46.3-million) contract to supply smallpox vaccine to a company whose chief recently donated £50,000 to the ruling Labour party.

The government denies that the donation influenced its decision and says the company, Powderject Pharmaceuticals of Oxford, is the only one that can supply the Lister strain of vaccine quickly enough. A spokesperson for the Department of Health says that it requested this strain of vaccine because it was the one used in Britain before the disease was eradicated in 1980.

Politicians on both sides have requested more information, and the opposition Conservative party has demanded an inquiry. One of the crucial questions, according to Labour's Ian Gibson, chairman of the House of Commons Science and Technology Select Committee, is why the government is only considering the Lister strain. The US government recently asked another British pharmaceutical company, Cambridge-based Acambis, to supply 92 million doses of a vaccine known as the New York City Board of Health strain. Virologists say that there is little difference between the two vaccines.



Cash injection: Powderject chief Paul Drayson made a donation to Britain's Labour party.

D. SILLITOE/GUARDIAN

Will the real Golgi please stand up

It was discovered more than a century ago, but cell biologists are still debating whether the Golgi complex is an autonomous entity. Erika Check profiles an organelle in identity crisis.

Ever since the Golgi complex was first described in 1898, this embattled cellular organelle has struggled to secure an identity for itself. Using his Nobel-winning method of staining cells with silver salts, the Italian biologist Camillo Golgi spotted that neurons contain a stack of flattened, membrane-bound sacs. But although the same structure is found in all nucleated cells, from humans to amoebae, naysayers argued for decades that it was merely an artefact of Golgi's staining technique.

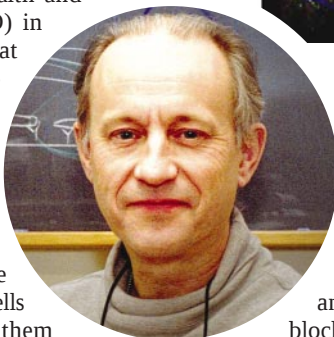
In the 1950s the electron microscope finally proved that the Golgi was no artefact. And the organelle gained further legitimacy when researchers revealed its crucial function: processing and packaging proteins for export from the cell.

But now, the Golgi's integrity is again in doubt. At issue is the question of whether it is a truly independent organelle, persisting through cell division in skeletal form and being rebuilt from this template. One camp of cell biologists, led by Graham Warren at Yale University in New Haven, Connecticut, subscribes to this view. But others, championed by Jennifer Lippincott-Schwartz at the National Institute of Child Health and Human Development (NICHD) in Bethesda, Maryland, argue that the Golgi is just a fleeting aggregation of proteins and lipid membrane that constantly assembles and disassembles.

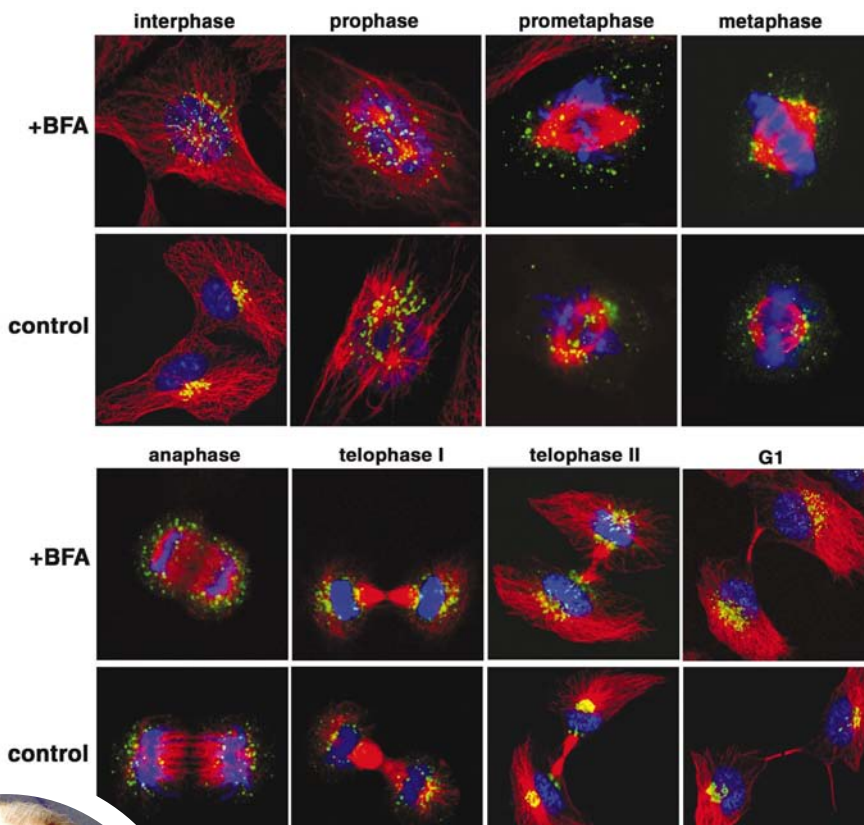
Dynamic cells

The Golgi question plays into a wider debate about how cells are built. The classic view is that cells make new parts by assembling them on static frameworks, rather as a skyscraper is built around a steel scaffold. But some biologists argue that many intracellular structures are constantly forming and disappearing in a flexible process that does not depend on underlying templates — like the dynamic equilibrium between condensation and evaporation in a billowing cloud. "This idea of dynamic self-organization is becoming more and more popular in many areas of cell biology," says Ben Glick, who studies the Golgi complex at the University of Chicago.

The Golgi's current identity crisis stems from the late 1980s, when Lippincott-



Graham Warren thinks that 'matrix' proteins (stained green, above) persist throughout cell division and act as templates for the Golgi's reassembly.



Schwartz was working in Richard Klausner's lab at the NICHD as a postdoc. She gave cells a shot of brefeldin A, an antibiotic from fungi that blocks the transport of proteins from the endoplasmic reticulum (ER) to the Golgi. The ER is another complex of intracellular membranes; it receives proteins directly from protein-building ribosomes attached to its surface. When transport from the ER to the Golgi was blocked, the Golgi disintegrated, and proteins associated with its membranes were rapidly redistributed to the ER. And when Lippincott-Schwartz removed the antibiotic, the Golgi reappeared¹.

To Lippincott-Schwartz and her colleagues, this indicated that the Golgi is constantly recycled to and from the ER in a dynamic equilibrium². "This was the first

crack in the theory that the Golgi is a stable, pre-existing entity," she says.

More recently, her group studied what happens to the Golgi during cell division, when its structure temporarily breaks down. By labelling Golgi proteins with green fluorescent protein, the researchers showed that these proteins fled to the ER. After cell division was complete, the ER spat out the Golgi proteins, and the structure rebuilt itself. But rebuilding could be prevented by expressing a mutant version of a gene called *Sar1*, which — like brefeldin A — blocks the transport of proteins from the ER³.

In the meantime, Warren and his colleagues had been looking more closely at cells disturbed by brefeldin A. Like Lippincott-Schwartz's team, they found that a dose of the antibiotic caused Golgi proteins to migrate to the ER. But not all of them. While working at the Imperial Cancer Research Fund's laboratories in London in the mid-1990s, Warren

showed that one protein, called GM130, stayed behind⁴. “This gave us the idea that such proteins might be the structure underlying the framework of the Golgi,” says Warren. “Taking an extreme view, you can argue that this is the Golgi apparatus itself.”

Although GM130 and other matrix proteins disperse through the cytoplasm when cells are dosed with brefeldin A, Warren has found that they still form Golgi-like structures in cells expressing a mutant *Sar1* gene, in which other Golgi proteins — the enzymes that process proteins destined for export from the cell — become confined to the ER⁵.

Tracking the template

In February this year, Warren and his colleagues reported on their efforts to track the matrix proteins during cell division. Again using brefeldin A or mutant *Sar1* to confine Golgi enzymes to the ER, they showed that matrix proteins were partitioned into the daughter cells in a manner reminiscent of the entire organelle. Using microscopic magnetic beads labelled with a fluorescent antibody that captures GP130, the researchers also found that this matrix protein remained distinct from the ER⁶.

This convinces Warren that matrix proteins are the template from which the Golgi is rebuilt after cell division. “Our take on this is that the Golgi is an independent organelle responsible for its own partitioning and the endoplasmic reticulum doesn’t play a part in this,” he says.

Lippincott-Schwartz interprets the results differently. She accepts that the matrix proteins stay separate from the main body of the ER in cells treated with brefeldin A. But she argues that the matrix proteins travel to a specialized part of the ER called its ‘exit site’, from which membranes pinch off and carry their cargo of proteins to the Golgi. It is this portion of the ER, Lippincott-Schwartz believes, that directs the Golgi’s reassembly.

She bases this claim on studies of dividing cells treated with brefeldin A, in which her team labelled the ER exit sites and matrix proteins using different fluorescent dyes. The two labels overlapped, and the labelled matrix proteins moved away from the Golgi before the other proteins in the organelle. This contradicts the idea that the matrix proteins stay behind after the Golgi unravels, Lippincott-Schwartz argues. Her team has also used a different mutant of *Sar1* that prevents the production of ER exit sites, and found that the Golgi completely fails to reform after treatment with brefeldin A⁷.

For now, the debate over whether the Golgi is an autonomous entity remains unresolved. Although Lippincott-Schwartz’s

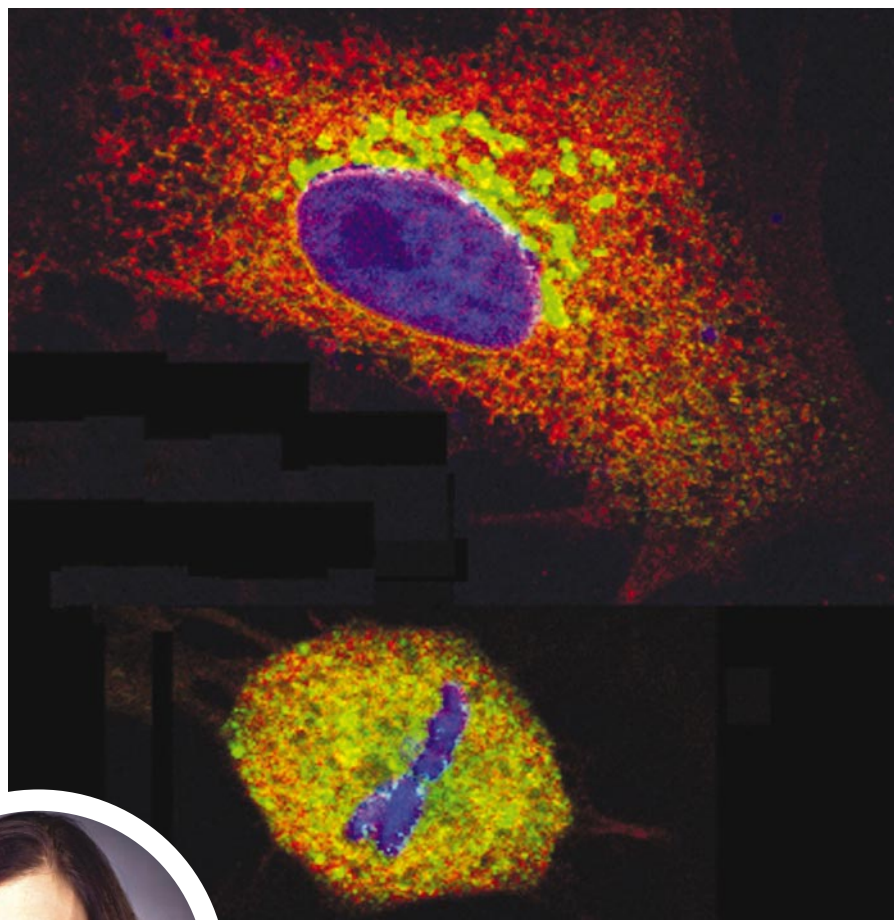


results suggest that the organelle is much more mutable than was once thought, other cell biologists say it is possible that we just do not know enough about the Golgi to be sure about what is going on when it reassembles. Perhaps the matrix proteins that direct the process are yet to be discovered, they suggest. “People are trying to define the Golgi based on four or five markers,” observes Vivek Malhotra of the University of California, San Diego. “What if there is a complex of proteins we don’t know about, and that is serving as a sort of nucleation site?”

Self-sufficiency

In addition, say some experts, we need to find out more about what happens to the Golgi during and after cell division, in the absence of any experimental disruption. If the Golgi reforms very quickly after cells divide, it would support the idea that it is being rebuilt from a residual template, rather than being recycled from the ER and assembling completely from scratch.

But the idea of dynamic self-organization has attractions that extend well beyond the Golgi. The reassembly of the nucleus from pre-existing skeletal structures might be



Mixing of Golgi (green) and endoplasmic reticulum (red) before a cell divides (bottom) makes Jennifer Lippincott-Schwartz question the Golgi’s autonomy.

the exception rather than the rule. If cellular structures were mostly assembled through dynamic and self-organizing protein interactions, it would help explain cells’ tremendous flexibility in responding to changes in their environment.

“Self-organization makes a lot of sense when you think about what a cell has to do,” says Tom Misteli, a cell biologist at the National Cancer Institute in Bethesda. “It allows a cell to be very stable, but on the other hand something terrible can happen to a cell at any moment, and it has to be able to respond to that.”

Seen in this light, the Golgi’s identity crisis seems less neurotic than noble. Far from being an isolated search for legitimacy, its resolution might provide fundamental insights into the way cells are built.

Erika Check is Nature’s Washington biomedical correspondent.

1. Lippincott-Schwartz, J., Yuan, L. C., Bonifacio, J. S. & Klausner, R. D. *Cell* **56**, 801–813 (1989).
2. Klausner, R. D., Donaldson, R. G. & Lippincott-Schwartz, J. *J. Cell Biol.* **116**, 1071–1081 (1992).
3. Zaal, K. J. M. et al. *Cell* **99**, 589–601 (1999).
4. Nakamura, N. et al. *J. Cell Biol.* **131**, 1715–1726 (1995).
5. Seemann, J., Jokitalo, E., Pypaert, M. & Warren, G. *Nature* **407**, 1022–1026 (2000).
6. Seemann, J., Pypaert, M., Taguchi, T., Malsam, J. & Warren, G. *Science* **295**, 848–851 (2002).
7. Ward, T. H., Polishchuk, R., Caplan, S., Hirschberg, K. & Lippincott-Schwartz, J. *J. Cell Biol.* **155**, 557–570 (2001).

The police state

On the surface, beehives and ant nests seem to be model societies, with each individual striving for the common good. But maintaining this social order sometimes calls for brutal tactics. John Whitfield reports.

Murder, torture and imprisonment — these are the standard tools of repressive regimes. But if you imagine that human societies have a monopoly on such tactics, think again. Social insects perfected the police state long before people got in on the act.

The workers in bee, ant and wasp colonies forgo their own reproduction to raise the brood of their queen. Such is a worker honeybee's devotion to the collective, she will eviscerate herself by stinging any animal rash enough to attack the hive — a selflessness that has long fascinated naturalists.

We now know that workers receive their evolutionary payoff through their relatedness to the monarch. As the queen's daughters, they share many genes with her other offspring. By caring for the regal brood, they ensure their own genetic inheritance.

But in any society, it is possible to prosper by breaking the rules. Worker honeybees do occasionally lay their own eggs. As well as being against the queen's interests, this puts them at odds with the rest of the workforce. And in the late 1980s, evolutionary ecologist Francis Ratnieks, now at the University of Sheffield, UK, showed that honeybee workers act as a 'police force' that cracks down on miscreants, eating their eggs¹. "Worker policing is a mechanism by which a society resolves its conflicts," says Ratnieks. "I think it's the best example of conflict resolution in nature."

Policing by egg-eating is common to all species of *Apis*, not just the familiar domestic honeybee, *A. mellifera*². And bees are not alone — the phenomenon has recently been observed in some species of wasp³ and in a group of ants⁴. "It may be very widespread," says Ratnieks.

But sometimes policing breaks down. Beehives occasionally dissolve into anarchy, as workers begin to reproduce *en masse*⁵. And a parasitic subspecies of honeybee is wreaking havoc in South Africa, destroying colonies by evading their police⁶. Evolutionary biologists hope that studying such face-offs between authoritarianism and anarchy will provide a deeper understanding of the balance between cooperation and selfishness



A honeybee kills an a renegade worker's egg, left.

at many levels, from genes within the genome to individuals in human societies.

Sociality has evolved about a dozen times in the Hymenoptera — bees, wasps and ants. This is probably linked to their unusual method of sex determination, called haplodiploidy. In the Hymenoptera, sexual reproduction always produces female offspring, which are diploid — they have two sets of chromosomes, just like humans of both sexes. Male hymenopterans, however, develop from unfertilized eggs, and carry only one set of chromosomes. These 'haploid' males produce sperm that are genetically identical. This is different from diploid animals, which shuffle

maternal and paternal chromosomes into their haploid sperm or eggs, giving infinite genetic possibilities.

Family ties

In human families, brothers and sisters each share half of their genes by common descent — their coefficient of relatedness (r), to use the technical term, is 0.5. The same is true for parents and offspring. For half-brothers or half-sisters, r is 0.25.

But haplodiploidy creates some bizarre family dynamics (see diagram, opposite). A mother's relatedness to her son is 0.5, but consider the same relationship from his perspective and r rises to 1.0. Sister-brother relationships are also unequal: from the sister's viewpoint, r is 0.25; from the brother's, it is 0.5. Mothers and daughters have a relatedness of 0.5, whichever way you look at it, and r for two sisters is always 0.75.

These final two figures are key, because they mean that a female can pursue her genetic interests more effectively by raising her mother's daughters rather than her own. This helps to explain why most females in social Hymenoptera do not reproduce. If the



Police supervisor: Francis Ratnieks was the first to describe honeybees' disciplinary tactics.

F. RATNIEKS

F. RATNIEKS

A. G. HART



An egg-laying worker bee (marked) is ignored...

queen is mother to the entire female workforce, they have little to gain by producing their own daughters.

In many species, workers are physically incapable of mating. And in honeybees, the monarchs exploit their subjects by mating with up to 20 males; workers are thus tending to young that are mostly half, not full, sisters, with a relatedness of 0.375 rather than 0.75.

Worker honeybees cannot rebel by mating, but they can lay unfertilized eggs that will develop into sons, with a relatedness of 0.5. But r once again explains why other workers destroy sneakily laid eggs. To most of the workforce, another worker's son is a half-nephew, with a relatedness of just 0.125, so the majority gain if workers' eggs are culled and everyone concentrates on raising the queen's brood. The result, as Ratnieks has found, is a police force so pervasive that it rivals the former East Germany's infamous Stasi.

God save the queen

At the top of the security apparatus sits the queen, producing pheromones that instruct the workers to carry on with the usual business of repression. Most loyally shut down their ovaries and keep watch for any subversives. In this way, the success of worker policing is conspicuous by the rarity of transgressors. Only about 3 of the 30,000 or so worker honeybees in a typical hive have functioning ovaries, showing the inequality of the struggle between a lone selfish worker and the rest of her caste. Although these rogue females can lay about 7% of a colony's male eggs, effective policing means that only 1 in 1,000 males has a worker for a mother⁷.

Indeed, workers' eggs usually get eaten within hours of being laid⁸, betrayed by their lack of a chemical badge unique to the queen's eggs⁸. Honeybee workers also seem to detect when their colleagues are about to break the rules, becoming aggressive to workers with active ovaries⁹. But as with human dictatorships, the death of the leader can trigger major societal changes. When the queen dies, and her pheromones no longer pervade the hive, the workers



...but this ponerine ant is assaulted by fellow workers as a punishment for attempting to reproduce.

hurl themselves into a frenzy of egg-laying¹⁰.

Honeybees are the most advanced insect police state so far discovered. Wasp societies are smaller and simpler, but give another elegant example of animal behaviour matching evolutionary theory. Again, it all comes down to r . In some species of vespine wasp, the queen mates with just one male, whereas in others she mates repeatedly. Where queens are monogamous, and all workers are therefore full sisters, the relatedness between a worker and a fellow worker's son rises to 0.375, compared with 0.125 for species in which the queen mates several times. In both cases, workers are related to the queen's sons by an r of 0.25.

So workers should tolerate one another's egg-laying if their queen mates just once — because 0.375 is greater than 0.25 — but clamp down on renegade egg-layers if the queen is more promiscuous. Sure enough, this is what Ratnieks and his Sheffield colleague Kevin Foster found when they compared different wasp species³. They even saw policing varying within species: in the saxon wasp (*Dolichovespula saxonica*), queens may

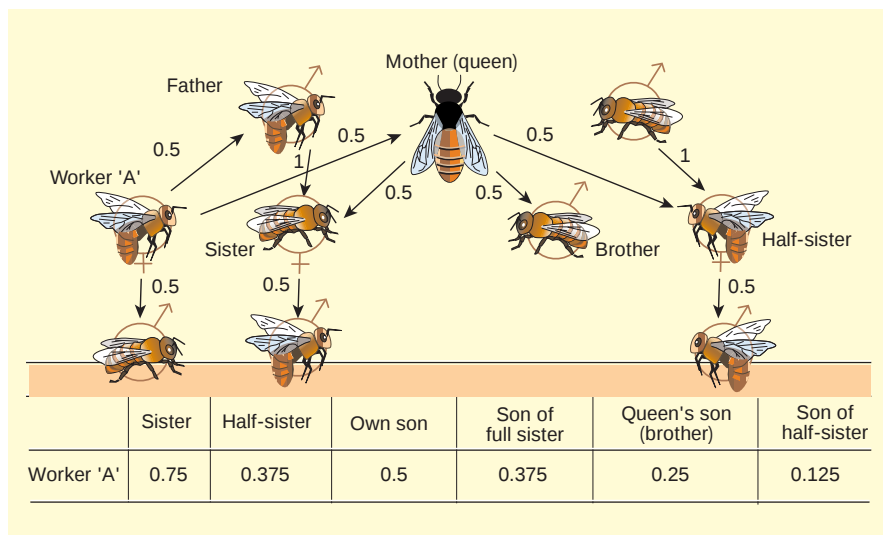
mate once or several times. Policing only occurs in colonies with promiscuous queens¹¹.

The fascist regime

In some ponerine ants, both social structure and policing are much less sophisticated. Living in colonies of 100 or fewer individuals in tropical forests, these ants have no rigid divide between queen and workers. Instead, all females have the potential to reproduce sexually. But only one or a few do, because small colonies cannot support many breeders.

Breeding females are called gamergates, and head a pecking order that has older workers at the bottom and younger workers — wannabe gamergates — in the middle. This hierarchy is enforced by both egg-eating and more brutal policing methods. If a young worker attempts to reproduce, she is spreadeagled by her fellows and kept immobilized for hours or even days. At the end of her sentence, the best she can hope for is a reduction in rank and loss of reproductive capability. Often she is mutilated or killed⁴.

But ponerine gamergates do not enforce the same iron control over their security appa-



Numbers game: to determine the degree of relatedness between a worker and her nest-mates, follow the arrows, multiplying the values on each route and adding the products.

▶ ratus as honeybee queens. From time to time, the ant police force stages a coup, deserting a failing ruler for a young pretender. "When the current reproductive is declining in fertility, workers can change allegiance," says Christian Peeters of the Pierre and Marie Curie University in Paris, who studies ponerine ants.

Even in honeybees, workers can overthrow their monarch's reproductive monopoly by activating their ovaries and tolerating one another's eggs⁵. But such anarchic hives are very rare. Ben Oldroyd of the University of Sydney, who studies the phenomenon, has sought out examples by placing adverts in beekeeping magazines. "We've had two reports in seven years of advertising," he says.

Oldroyd suspects that this is because honeybee anarchy requires a mutational double whammy: worker bees must ignore the pheromones that suppress their ovaries, and must also produce eggs that evade policing, probably by counterfeiting the queen's chemical signal. Oldroyd's team has shown that ovary activation and egg-faking are under separate genetic control: they have managed to breed both egg-laying workers whose eggs are detected and destroyed by their fellows, and workers that do not lay eggs, but do not destroy workers' eggs transplanted into the colony¹².

Breeding for anarchy

Naturally anarchic hives contain a few dozen laying workers, and seem to function well. But by selective breeding, Oldroyd and his colleagues have got the egg-layers up to about 40% of the workforce. In these hives, breeding workers neglect their chores and the hives become decadent to the point of collapse². "They can barely feed themselves, and they do weird things like trying to raise queens out of male larvae," Oldroyd says.

These selectively bred anarchist hives have some other odd features. Normal workers transplanted into them may start laying¹³, which suggests that the queen's pheromones are weaker than normal. Most strangely, taking the queen out of an anarchic colony — which triggers worker egg-laying in normal hives — causes anarchic workers' eggs to lose their deceptive properties. They are eaten if transplanted into a normal hive, says Oldroyd.

Oldroyd and his team are trying to identify the genes that influence anarchic behaviour.



Point of order: Ben Oldroyd has bred anarchic bee colonies, but this is very rare in nature.

If they can find them, it will provide a long-awaited mechanism for theories about animal conflict and cooperation. "Anarchy genes are genes for selfishness, and anti-anarchy genes are genes for altruism," says Oldroyd.

A honeybee hive's policing can also be subverted from the outside. In January this year⁶, Ratnieks's team described studies of a parasitic form of the Cape honeybee (*A. m. capensis*), a freeloader that over the past decade has cut a swathe through hives of the African honeybee (*A. m. scutellata*) in northern South Africa. The parasitic Cape workers are like Oldroyd's anarchists in that they do little work. But instead of laying haploid male eggs, they produce diploid female eggs that are genetic copies of themselves¹⁴.

Ratnieks and his colleagues have shown that the African workers cannot distinguish these eggs from those of their own queen — presumably because the Cape bees use some sort of chemical counterfeit. As a result, the invaders' numbers grow so quickly that the colony is destroyed within months. The Cape bees then move on to another hive.

Formerly confined to the southernmost part of South Africa, Cape honeybees were probably spread to the hotter north of the country, which is favoured by *A. m. scutellata*, by South African beekeepers who move their hives around to catch the blooming seasons of different flowers. This nomadic beekeeping style is still aiding the parasite's spread.

The parasitic Cape bee now plagues an area of South Africa about 400 by 200 kilo-

metres in size, stretching from Pretoria to the border with Mozambique. The parasite's ability to produce exact clones of itself, keeping its clique of selfish genes together, makes it a formidable fifth-columnist. "We spread the net as far as we could, and everything we've caught has been the offspring of a single worker," says Per Kryger of the University of Pretoria, who is monitoring the problem.

Bees not for keeps

"Many beekeepers are losing 100% of their colonies each year," Kryger says. "The only reason they're still in business is that they collect swarms from the wild." Kryger estimates that 50,000–100,000 wild *A. m. scutellata* swarms per year are lost to domestication — and subsequent parasitism. If the trend continues, he says, the local survival of the African honeybee will be threatened.

Kryger is trying to work out whether the parasitic form is also common in the Cape bee's home range. He suspects that *A. m. capensis* may have evolved countermeasures to deal with its seditious element. Exactly what these might be remains unclear — normal Cape honeybees show policing behaviour, but this is not especially pronounced¹⁵.

Studies of such conflicts may also yield insights into other aspects of biology, and even the social sciences. Laurence Hurst, an evolutionary geneticist at the University of Bath, UK, draws parallels between worker policing and the mechanisms that genomes have evolved to prevent individual genes from inserting themselves into a disproportionate number of sperm or eggs.

"Worker policing is a nice case history for examining the conflicts present in any social structure," Hurst says. He believes that it may be possible to derive theoretical models to capture accurately the dynamics of conflict in everything from genomes, through social insects, to human society. "There are important differences between different systems, but there are also abstract features that unite them. It isn't just a metaphor," he says. ■

John Whitfield works in Nature's news syndication team.

1. Ratnieks, F. L. W. & Visscher, P. K. *Nature* **342**, 796–797 (1989).
2. Barron, A. W., Oldroyd, B. P. & Ratnieks, F. L. W. *Behav. Ecol. Sociobiol.* **50**, 199–208 (2001).
3. Foster, K. R. & Ratnieks, F. L. W. *Behav. Ecol. Sociobiol.* **50**, 1–8 (2001).
4. Monnin, T. & Ratnieks, F. L. W. *Behav. Ecol. Sociobiol.* **50**, 97–108 (2001).
5. Oldroyd, B. P., Smolenski, A. J., Cornuet, J. M. & Crozier, R. H. *Nature* **371**, 749 (1994).
6. Martin, S. J., Beekman, M., Wossler, T. C. & Ratnieks, F. L. W. *Nature* **415**, 163–165 (2002).
7. Visscher, P. K. *Behav. Ecol. Sociobiol.* **39**, 237–244 (1996).
8. Ratnieks, F. L. W. *J. Apic. Res.* **34**, 31–37 (1995).
9. Visscher, P. K. & Dukas, R. *Anim. Behav.* **49**, 542–544 (1995).
10. Page, R. E. & Erickson, E. H. *Behav. Ecol. Sociobiol.* **23**, 117–126 (1988).
11. Foster, K. R. & Ratnieks, F. L. W. *Nature* **407**, 692–693 (2000).
12. Oldroyd, B. P. & Osborne, K. E. *Proc. R. Soc. Lond. B.* **266**, 1335–1339 (1999).
13. Barron, A. B. & Oldroyd, B. P. *Behav. Ecol. Sociobiol.* **49**, 214–219 (2001).
14. Verma, S. & Ruttner, F. *Apidologie* **14**, 41–57 (1983).
15. Moritz, R. F. A., Kryger, P. & Allsopp, M. H. *Behaviour* **136**, 1079–1092 (1999).



Insect invasion: an African honeybee (right) unwittingly feeds one of an army of parasitic Cape bees.

'Practical autonomy' entitles some animals to rights

Basic rights stem from the ability to desire, to act intentionally and to have a sense of self.

Sir—In your Opinion article "Rights, wrongs and ignorance" (*Nature* **416**, 351; 2002), you urge scientists and lawyers to challenge the arguments for the basic rights of great apes and other non-human animals that I made in my books *Rattling the Cage* (Profile Books/Perseus; 2000) and the forthcoming *Drawing the Line* (Perseus; 2002). You say that some cognitive scientists will challenge my interpretation of the scientific evidence. But others will not. Those who do will find themselves boxing not with me, but with many eminent primatologists and cognitive scientists, for it is their interpretations I adopt.

I am more qualified to argue about law. You refer to those who claim that only beings able to assert rights can have them, so therefore neither young children nor non-human animals can have rights, but should instead be protected. Presumably you would extend this argument to any human beings unable to assert their rights. Yet even the philosopher Carl Wellman, a major proponent of that idea, calls this extrapolation monstrous. Unlike claim rights, such immunity rights as bodily

integrity and freedom from slavery can belong to human children, infants, the very retarded, the profoundly senile and the insane. And thank goodness for that, as it is the weakest humans who are in most need of legal protection from exploitation.

You challenge my argument that the precautionary principle, a staple of environmental law, should encourage judges and legislators to recognize basic rights for some non-human animals, and say that the "immense benefits of biomedical experimentation for human health" should be factored in. This misunderstands both my argument for the basic liberty rights of non-human animals and the precautionary principle itself.

I say that a minimum level of autonomy—the abilities to desire, to act intentionally and to have some sense of self, whatever the species—is sufficient to justify the basic legal right to bodily integrity. I call this level "practical autonomy" and maintain that a creature who demonstrates it, whether an adult chimpanzee, bonobo, gorilla, orang-utan, Atlantic bottlenosed dolphin or human, is

entitled to this basic legal right. One may reject my premise. But I argue that this rejection on utilitarian grounds undermines the foundation for most basic human rights. We don't enslave people or deprive them of their right to bodily integrity because we think we will benefit if we do.

The precautionary principle comes into play when one is not as sure as one would like that a being has practical autonomy. The strength of that being's claim to bodily integrity turns on how uncertain we are. I argue against both stringent and loose applications of the precautionary principle to the question of which non-human animals are entitled to basic legal rights.

Under the moderate application I urge, African grey parrots and African elephants are entitled to basic rights. Based on what is known, dogs and honeybees are not. But who knows what exciting breakthroughs tomorrow's research may bring?

Steve Wise

Center for the Expansion of Fundamental Rights, Inc., 896 Beacon Street, Boston, Massachusetts 02215, USA

Cancer centre didn't shoot the messenger

Sir—In your Opinion article "Media studies for scientists" (*Nature* **416**, 461; 2002), you imply that scientists at the Fred Hutchinson Cancer Research Center failed to provide information to *The Seattle Times* and then cried "bad journalism". You advise that "Rather than shooting the messengers, scientists should take them to one side and give them the real story".

Researchers at the Fred Hutchinson Cancer Research Center enjoy a good relationship with the media, frequently spending considerable time discussing and explaining our science to reporters. If you had checked with us you would have discovered that, in this case as well, we logged many hours of one-to-one interviews between researchers at the centre and the reporters for *The Seattle Times*.

We will continue these efforts to work with the news media on our campaign to inform the general public about the importance of clinical trials. We hope that reporters will invest the time to explore and discuss fully the subject about which they are writing.

It is also noteworthy that the accusation

of bad journalism came not from a scientist but from a journalist for *The Wall Street Journal*.

Lee Hartwell

Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109-1024, USA

A theory can be falsified or tested. Faith cannot

Sir—Your news story "Evolution critics seek role for unseen hand in education" (*Nature* **416**, 250; 2002) identifies an important issue in science teaching: that of the way science is taught versus the concepts that are taught as science. The proponents of 'intelligent design' correctly point out the value of challenging scientific theory. Such challenges are an essential component of science if theories are not to become dogma. Science curricula the world over should embrace this principle, and science should be taught as a dynamic field in which new ideas are continually being tested and revised.

Scientific progress has been made by examining many wondrous and fantastic ideas, but educators must make it

abundantly clear that not all fantastic ideas represent sound science. This is central in determining the subject matter of science curricula. The longstanding and powerful attempt to integrate creationism into science classes is an excellent example of politics blurring the boundary of the subject.

Creationists argue that since evolution is not completely proven, it is important to present a competing theory, to let students decide for themselves. Scientists oppose such ideas for one simple reason: creationism, in any form, is not science. By including a supernatural being, or an 'intelligent design' concept, creationists exclude their ideas from the domain of science because these are neither testable nor falsifiable—two criteria that are crucial to scientific method.

Any alternative to evolutionary theory must follow the methods of science, being based on testable, falsifiable hypotheses. Perhaps one way in which creationism can be used constructively in science education would be as a tool to illustrate how not to carry out science.

Mark W. Silby

Center for Adaptation Genetics and Drug Resistance, Department of Molecular Biology and Microbiology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, Massachusetts 02111, USA

A few words about evolution

Building a hierarchical framework on the foundations of darwinism.

The Structure of Evolutionary Theory

by Stephen Jay Gould

Belknap: 2002. 1,464 pp. \$39.95, £27.50, 45.90 euros

David B. Wake

We live in a darwinian world. For many evolutionary biologists the ever-extending grand evolutionary synthesis promulgated in the 1930s and 1940s offers adequate explanations and serves as a consistent generator of testable hypotheses. For Stephen Jay Gould it is but a foundation, wholly inadequate to be a framework within which to build. Evolution proceeds in a hierarchical manner, so proper analysis requires a hierarchical approach, Gould believes.

Modern darwinians, being largely reductionist in approach — while paying some lip-service to the possibility of constraints, the reality of mass extinctions and the like — view evolution in general as microevolution plus lots of time and some contingency. Gould strongly rejects this perspective. His views are well known, but they have had surprisingly little effect on some of the core disciplines in evolutionary biology, such as population, and quantitative and evolutionary genetics, whose practitioners see themselves as defenders and extenders of the darwinian message. In *The Structure of Evolutionary Theory* Gould makes a direct, intellectually rigorous argument that a hierarchical perspective would not only enrich modern darwinism, but that it is essential. This book is a manifesto for a new kind of evolutionary biology, one that makes full use of many kinds of knowledge as well as diverse 'ways of seeing'.

This large book demands attentive reading. Gould uses many more words than most writers and the book appears not to have been edited at all. Those who enjoy Gould's erudition and style will find this to be a stimulating and illuminating book; it will infuriate others. There are intellectual wars in progress in the evolutionary-biology community, but most researchers seem to care little about battles engaged around issues such as hierarchical versus reductionist approaches, and structuralist versus functionalist versus historical perspectives. I believe that these are important issues, and that in the long run evolutionary biology will have to confront them. Gould's book, an intellectual *tour de force* and a work of serious scholarship, will be a permanent factor in the struggle to understand how life has evolved.

Gould knows the book is long and daunt-



ing and has supplied readers with a kind of abstract, itself 89 pages long. In this opening chapter he previews, but cannot really encapsulate, the many deeply analytical historical and philosophical treatments to come. This is a long and circuitous argument that new knowledge and new macroevolutionary explanations have been so substantial that a full exposition of evolutionary theory within the domain of darwinian logic "must be construed as basically different from the canonical theory of natural selection, rather than simply extended". This point is made repeatedly throughout the book, lest anyone miss it.

The strategy is to examine the historical roots of evolutionary theory in great detail and to show how earlier workers, even Darwin (but excluding Wallace, who remained the ultimate pan-selectionist), were forced to resort to hierarchical thinking at some point in their intellectual journeys. The book is a veritable history of evolutionary biology, but not all of it, of course. Gould sets the agenda and discusses what he wishes from his point of view: that of a strong advocate for an expanded and greatly modified evolutionary theory.

All Gould's familiar arguments are here, with emphasis on levels of selection, punctu-

ated equilibrium, and their contribution to the development of Gould's hierarchical perspective. His points are forcefully made, but the arguments of only some of his detractors are considered. Gould makes a convincing case for the "hardening" of the adaptationist perspective during, and especially at the end of, the grand evolutionary synthesis in the middle of the twentieth century. He charges that evolutionists became almost blindly channelled. Selectionists will complain that their position is oversimplified.

In a time when we are in something of an intellectual revolution in our understanding of species and species formation, Gould is unrelentingly stubborn in his insistence that speciation is a big deal, that species are individuals, and that species selection is the critical keystone of a hierarchical approach. Close study of species formation has made me question the reality of species as bounded entities, and I am wary of the perspective that species are individuals. I find clade selection an attractive and more general alternative to species selection, but Gould will not give ground. Many of us are attracted to aspects of punctuated equilibrium, and the concept has stimulated much research over the past 30 years. But the insistence that it is a theory of the deployment of species over time and

space makes many uneasy and leads to doubts that it is the core concept for the new hierarchical theory of evolution.

I thoroughly enjoyed many parts of the book, for example the long and rewarding chapter on the evolution of development with its strong focus on the positive evolutionary effects of constraints, and the concluding chapter that deals extensively with spandrels. The structuralist message of these chapters is strong and clear, but it is integrated with functionalist and historical perspectives as well. Parallel evolution is given the attention I believe it merits, and Gould does an admirable job of showing why it is important. A fan of *Hox* genes (he loves the word 'hoxology'), Gould also shows the importance of developmental genetics for his hierarchical perspective and for understanding the evolution of form in organisms.

Gould, the most widely known evolutionist of our time, has remained active in research while at the same time communicating his evolutionary message to the public through his prolific writings and appearances. This book demonstrates that he is not just a popularizer, but a major intellectual force in his discipline. Yet one can predict a strong negative reaction because the book is annoyingly self-congratulatory and self-serving. But what can he do? He is certain that he is right! A strength, and at the same time a weakness, of this book is that the author's powerful personality emerges on nearly every page. The important messages of this book are appreciated most fully when the reader accepts and enjoys the idiosyncrasies of this extraordinary man. ■

David B. Wake is at the Museum of Vertebrate Zoology, University of California, Berkeley, California 94720-3160, USA.

The nature of nature's economy

Human Well-Being and the Natural Environment

by Partha Dasgupta

Oxford University Press: 2001. £25

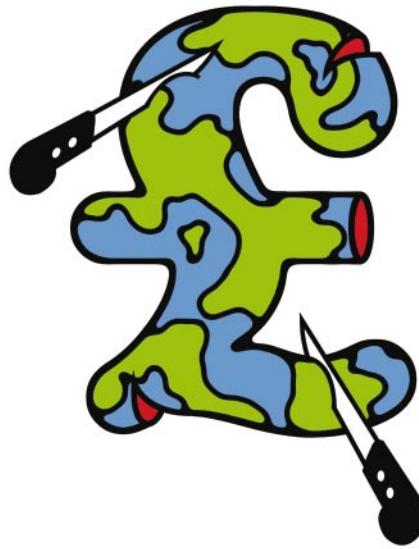
The New Economy of Nature: The Quest to Make Conservation Profitable

by Gretchen Daily & Katherine Ellison

Island Press: 2002. £19.50

Norman Myers

We overexploit our environments and we underutilize them. We convert forests into newspapers with disregard for their many other outputs such as watershed functions and climate regulation. We drain wetlands for housing to the detriment of their role in flood control and as species habitats. Yet it is the abundant 'ecosystem services' that often



offer more to society than limited products with limited benefits. But everybody's benefits tend to be nobody's business as nobody owns the ecosystems services, so nobody has an incentive to protect them. Everybody's atmosphere becomes everybody's garbage bin for pollutants, and everybody's river becomes everybody's sewer.

Let us welcome, then, these two books, which show us how to get on top of the problem before it gets on top of us. *Human Well-Being and the Natural Environment*, by a past president of the British and European Economics Associations, focuses on the contributions of ecosystem services to the quality of life, with emphasis on impoverished communities of developing countries. The author is concerned that even though the natural environment is central to our lives, it is "absent from national accounts, from quality-of-life indices, and, more generally, from official development economics". Hence, we "must stop viewing the environment as an amenity, a luxury the poor can't afford". In fact, Dasgupta argues that the environment is often the greatest asset for poor families because they have few alternatives for income if it fails.

The book ranges over such disparate topics as well-being (both personal and social), GNP as a measure of well-being (deeply inaccurate and even misleading as it is), the Human Development Index (the United Nations' measure of social and economic advancement), poverty and consumption, population pressures, technological change, markets and other economic institutions, common property resources, economic efficiency versus social equity, intergenerational conflicts, human rights and civil rights. All are accorded a firm theoretical treatment. The book is solidly interdisciplinary, covering not only economics but ecology, anthropology, political science, philosophy and ethics. It should make a first-rate reader for the World Summit on

Sustainable Development in Johannesburg next August.

While it gives a comprehensive account of what makes up human well-being, the book frequently diverges from certain conventional views. "Countries that would be regarded as having performed well if judged on the basis of such indices as GNP per head or the Human Development Index are found to have grown poorer... the poorest countries have 'developed' by depleting natural capital relative to their high population growth rates." Indeed, Dasgupta deeply disagrees with those many economists who assert that high population growth in poor countries has not slowed their development. Rather, he sees the problem lying with an interactive mix of "high fertility, poverty, malnutrition, illiteracy, and degradation of local natural resource bases, all of which feed on one another cumulatively". Decline of ecosystem services can often serve as a crucial catalyst of deficiencies in other sectors. In other ways, too, the book's message contrasts sharply with much contemporary thinking on economic development.

Parts of the book will prove too dense and technical for those with expertise outside economics. But Dasgupta outlines his theme in an introductory summary, a sort of roadmap for the many analytic pathways ahead. Each of the book's five parts is preceded by a prologue, further enabling the lay reader to grasp the guts of the discussion in short order. Moreover, Dasgupta frequently engages in happy phraseology that leavens the otherwise heavy text. For instance: "The present is the past's future, [and] the future has an unnerving habit of becoming the present." These stylistic flourishes help to make the book more accessible, which is all the more pertinent in that it addresses factors central to the human condition writ large. Indeed, Dasgupta is at pains to talk in terms more general than economics-speak, and algebra is confined to a set of appendices. For the specialist, there are over 500 references.

The New Economy of Nature demonstrates that ecosystem services, although nobody's business in established practice, can become profitable business for "green gold prospectors". A good number of innovative entrepreneurs, both private and public, are making big money from these services. One well-known example is New York City, which depends for its water on a watershed, the Catskill Mountains, 125 miles away. The water supply could be maintained either through a water-filtration plant costing \$6-8 billion, plus annual maintenance expenses of \$300-500 million, or through an investment of \$1.5 billion to restore the watershed. No fewer than 140 American municipalities are now calculating the costs of protecting watersheds compared with the costs of building water-purification

plants. They thus demonstrate that eco-systems are to be viewed as capital assets, supplying both services and savings.

There are a host of similar initiatives around the world. The authors describe in detail how Australia is tackling its farmland salinity crisis through its Murray–Darling Basin Initiative, which, with an investment of US\$1.4 billion over the first three years, plus a major tree-planting effort, covers 360,000 square miles and accounts for 40% of the country's agricultural production. The United States offers cash payments for a range of environmentally supportive farming practices. Similarly, Costa Rica supplies payments to farmers who protect ecosystem assets on their land, with compensation based on the ecosystem services provided. The book presents additional breakthroughs in Vancouver, Adelaide and Rio de Janeiro, as well as broad-scope efforts such as tradeable pollution permits for sulphur and carbon dioxide, and safeguard measures for wild pollinators that support one-third of our food crop harvests. The authors emphasize that there are now hundreds of such enterprises, all showing that conservation need not always cost money, rather, it can save it in huge amounts.

In essence, the strategy counters a basic environmental problem: that goods and services without price are usually viewed as without value. As a result, to cite Richard Sandor, a financial innovator in Chicago who has been devising ways to mass market a number of environmental commodities: "Nature [without prices] becomes like an all-you-can-eat buffet" — and few people don't overeat at a buffet. A further result, as the authors document with illuminating detail, is that the "importance of ecosystem services is often widely appreciated only upon their loss". Fortunately, the green gold entrepreneurs are "establishing a truly new economy of Nature, with focus on the self interest factor that makes the strategy self sustaining". The challenge is to change the rules of the game so as to produce new incentives for environmental protection, readily available through the profit motive.

The two books are exemplary expositions of the environment's role in fostering socio-economic advance as part of human well-being. Each in its way is enlightening from start to finish. ■

Norman Myers is Honorary Visiting Fellow, Green College, Oxford University, Upper Meadow, Old Road, Oxford OX3 8SZ, UK.

More on the environment

Cradle to Cradle: Remaking the Way We Make Things

by William McDonough & Michael Braungart
North Point Press, \$25

Eco-economy: Building an Economy for the Earth

by Lester R. Brown
Earthscan, £17.99

State of the World 2002

by The Worldwatch Institute
Earthscan, £12.95 (UK); W. W. Norton \$29.95 (hbk), \$15.95 (pbk) (US)

Guide to Sustainable Development and Environmental Policy

edited by Natalia Mirovitskaya & William Ascher
Duke University Press, \$89.95 (hbk), \$34.95 (pbk)

Building machines more like humans

Flesh and Machines: How Robots Will Change Us

by Rodney A. Brooks
Pantheon: 2002. 288 pp. \$26

The Body Electric: An Anatomy of the New Bionic Senses

by James Geary
Weidenfeld & Nicolson: 2002. 256 pp. £20
Igor Aleksander

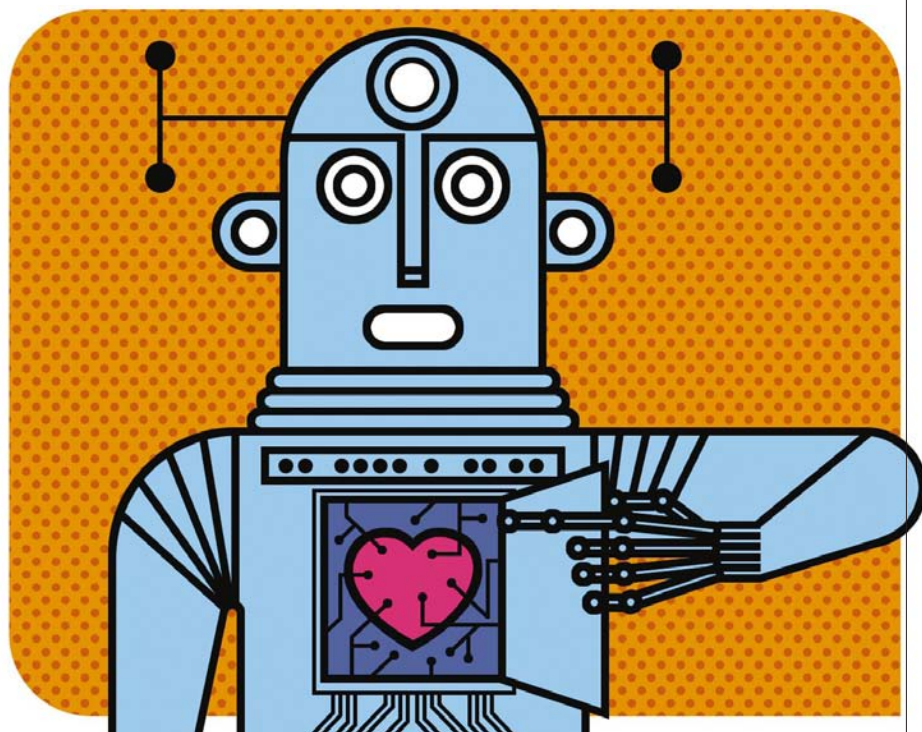
As robots are made to be more like humans, and human organs are replaced by mechanical artefacts, will humans cease to be more 'special' than machines? A hurried reading of the dust jackets of these two books may give the false impression that they are ill-informed texts written to create fears about advancing bionic aliens mischievously created in frankensteinian laboratories. Actually, nothing could be further from the minds of

the authors. Both books cover serious scientific and philosophical issues in the design of humanoid robots and the replacement or supplementation of organs.

Rodney Brooks, an influential roboticist from the Massachusetts Institute of Technology, has been building what he calls 'artificial creatures' for some years. In *Flesh and Machines* he writes autobiographically of his flight from 'representational' artificial intelligence, in which the programmer endows the robot with a full description of its environment so that it might act intelligently. Brooks now takes an alternative 'situated' approach in which the robot is equipped with ways of reacting appropriately to its environment without the burden of referring all its actions to pre-digested models supplied by the programmer. Indeed, one of his early models was an insect-like hexapod called Genghis that reacted to the heat emitted by mammals in its vicinity. Later systems include the celebrated Cog, a waist-up robot that moves its limbs and eyes appropriately in the presence of a human. Another machine, Kismet, responds to humans with appealing facial gestures.

An evident weakness in this approach is that Brooks may need to return to some inner representation to introduce the contemplation that humans use a lot of the time, before reacting to their environments. This does not need to be done through pre-programming but in a brain-like fashion through a rapid evolution of, and learning in, an artificial nervous system. Brooks argues that this is just an elaboration of what he has achieved already, but this remains to be seen.

The book moves from an autobiographical to a predictive style in which visions of



robots range widely from helpful house servants to agents on planetary missions. But my greatest enjoyment came from the end of the book, which is more philosophical. Brooks' philosophy is blunt: abilities that some hold dear as being specially human (for example, emotions and consciousness) are available to machines simply because humans are machines, albeit immensely complex ones.

He prepares a robust response to potential attacks by detractors who believe that the 'specialness' of human consciousness either lies in an as yet uncomprehended presence of quantum physical action in the brain, or is not available to physics at all. He accuses some philosophers of simply being afraid to contemplate the possibility of a conscious machine. According to Brooks, the adversaries will have to prove their points or admit that humans can be understood from physical control principles. Despite this, Brooks has his own little *je ne sais quoi* which separates the biological from the artefactual. The 'juice', as he calls it, that is missing is a proper mathematical understanding of complex biological mechanisms. He does not see this as an insurmountable difficulty, but rather as an achievable and exciting target for science.

The second book, *The Body Electric*, is a clear and readable report of the way that human sensory systems are being replaced or supplemented by technological artefacts studied in a variety of research laboratories across the world. Journalist James Geary tells an exciting story not only by reporting his in-depth interviews with key researchers, but also by asking the beneficiaries of some prostheses what this has done for them. People who have seen, heard and learned to speak for the first time are the best ambassadors for this science — their accounts detract from any false impression that humanity is being insidiously invaded by the silicon chip.

Inevitably, reportage can fall prey to some odd views held by those who are being interviewed. A British engineer, for example, insists that a radio chip he had implanted under the skin of his forearm to open doors and cause his computer to greet him is a major step towards the automatic transmission of emotions and, eventually, thoughts, between humans. Here the scientific challenges of encoding thought and emotion distributed over billions of brain cells are suspended, and would a chip in the pocket not have opened the door just as well?

By contrast, the rest of the book presents real scientific work on vision, hearing, touch and even taste. The question is always the same: how can a sensory faculty be restored to an individual who may have lost it? Often this needs an enormous amount of ingenuity, mainly because the human brain is adept at integrating the input from many sensory

modalities. For example, in order to 'see', the brain integrates signals from the retina with signals from the oculomotor system that moves the eyes. The key question is whether what's left of the brain's apparatus when parts go missing will be able to accept and adapt to the retinal implant in order to bring to the perceptual mind some sense of seeing. Geary answers this by sharing the delight of a 63-year-old Belgian woman who, after having been blind for over 20 years, began to see dots of colours when her optic nerve was stimulated by the output of a small video camera.

Having dealt with the senses, Geary approaches the problem of artificial minds. He talks to some who insist that they can create a mind just by throwing neural networks together and letting the mind emerge, as well as to others who have more principled approaches along the lines of the work reviewed by Brooks. Geary, rightly, just reports, and leaves it to the reader to spot the outliers. ■

Igor Aleksander is in the Intelligent and Interactive Systems Group, Department of Electrical and Electronic Engineering, Imperial College of Science Technology and Medicine, London SW7 2BT, UK.

Putting evolution in context

The Alfred Russel Wallace Reader: A Selection of Writings from the Field

edited by Jane Camerini

Johns Hopkins University Press: 2002. 248 pp. \$42.50, £29.50 (hbk); \$18.95, £13 (pbk)

Infinite Tropics: An Alfred Russel Wallace Anthology

edited by Andrew Berry

Verso: 2002. 320 pp. £19, \$27

Charlotte Sleight

Charles Darwin kept many of his opinions, especially on politics and religion, under his hat. As a result, biologists, historians, marxists, keynesians, atheists and fundamentalists have all enjoyed debating ever since what Darwin 'really' meant. This kind of parlour game is no fun with Alfred Russel Wallace. A far more colourful and opinionated Victorian, he promulgated his beliefs to anyone who would listen. Ranging through socialism, geography, exobiology, phylogeny, female suffrage, natural selection, land reform and spiritualism, his sheer breadth of interests creates a challenge for the anthologist: how to group his writings to give the truest impression of his life and work.

This is a healthy challenge. Darwin's silence on contentious issues has obstructed a truly historicized understanding of his natural history, suggesting to some that the



theory of evolution may be understood (could, even, have been created) in isolation from its social milieu. Wallace allows no such facile judgements. How should we appreciate the remarkable practical and theoretical achievements of this man, knowing that he promoted table-tapping and utopian community with equal seriousness and fervour?

Stephen Jay Gould, in his preface to Andrew Berry's *Infinite Tropics*, appears somewhat unnerved by this paradox, referring uncomfortably to Wallace's "promiscuous... outpourings" of non-scientific writing. Berry responds to the challenge by grouping Wallace's writings thematically, while recognizing that this artificially categorizes his interdisciplinary thought.

Jane Camerini's solution for her slimmer collection is to gather Wallace's writings chronologically. Each section is introduced by a succinct and historically sensitive essay, locating Wallace in his social and political context. She is at pains to show that his spiritualism and socialism were not intellectually anomalous, but were interwoven with his motives and methods for natural history. Where Berry leads the reader from extract to extract, patiently glossing each development in Wallace's thought, Camerini is more inclined to let Wallace's writing speak for itself, allowing the variety of material to suggest its own connections.

One interesting thread thus highlighted is the significance of land in Wallace's life and thought. Time and again it recurs in his writings: his youthful land surveys in Wales; his records of the relationship between the Welsh farming community and the land; the sense of injustice with which he describes the effects of the General Enclosure Act of 1845. In the same period, Wallace records his frustration that a botany book contained no information about plant distribution. Soon after, we see his punctilious mapping of the Amazon; his vital evolutionary connection drawn between phylogeny and distribution; and his socialist writing for the Land Nationalisation Society. Towards the end of his life he was critical of US land use. We also read his children's recollections of their mutual chagrin over

their father's disregard for 'No trespassing' signs during family walks, and the consequent confrontations with farmers and land-owners. This is an instance of how Wallace's science (in this case, what we now refer to as biogeography) was inextricably embedded in his experiences, and within social and political issues. These cannot be removed to leave the 'bare ideas' behind — they simply would not have existed without them.

Wallace's 1858 paper 'On the tendency of varieties to depart infinitely from the original type' was famously read at the same Linnean Society meeting as Darwin's hastily penned description of natural selection, in a gentlemanly resolution of the question of priority. *Infinite Tropics* contains more technical papers than *The Alfred Russel Wallace Reader*, but this one is rightly central to both collections. A close reading of Wallace's paper reveals two interesting differences in emphasis between it and Darwin's writing of the same era.

First, Wallace stressed competition in relation to the environment (whether organic or inorganic) and between species, rather than the interspecific competition, or "ten thousand wedges", which forms a major part of Darwin's *On The Origin of Species* and has retrospectively been defined as its crucial argument. In this respect, Wallace is more responsible than Darwin for the layperson's understanding of evolutionary factors — a struggle against predators rather than against one's fellow-species.

The second difference is that Wallace emphasizes the distinction between domestic and natural varieties. The latter are defined for him by the organism's need for the variant characteristic, its competitive advantage against other varieties, and (in 1858 at least) its strengthening through use. Together these factors produced an irreversible directionality in the genesis of new varieties in nature. Darwin, meanwhile, although he agreed with all these points individually, preferred to stress the similarities between natural and domestic variants in the construction of his argument. Readers were prepared for the idea that nature might select by comparison to the acts of a pigeon breeder, amongst other homely examples. Here, Darwin's version subtly but powerfully altered the reception of his argument, making selection more anthropomorphic and less environmental — less a product of the land, as Wallace would have had it.

The geologist Charles Lyell, the botanist Joseph Hooker and Darwin all failed to pick up on these differences while discussing the presentation of Darwin and Wallace's work as simultaneous discovery. This corroborates the arguments of recent historians (Gillian Beer, Adrian Desmond, James Moore and Robert Young) that Darwin's ideas were by no means as clear-cut as the authors of the 'new synthesis' — and

biology textbooks — would have us believe. Wallace's 'problematic' interests and perspectives are revealed by these excellent anthologies (especially Camerini's) to be key to understanding the mid-nineteenth-century debates about evolution in their true cultural complexity. ■

Charlotte Sleigh is at the Centre for History and Cultural Studies of Science, Rutherford College, University of Kent at Canterbury, Canterbury CT2 7NX, UK.

I am loved, therefore I think

The Cradle of Thought: Exploring the Origins of Thinking

by Peter Hobson

Pan Macmillan: 2002. 291 pp. £20

Simon Baron-Cohen

How does the mind grow? This is a question that has exercised some of the best minds over the centuries, and Peter Hobson joins this laudable effort to try to answer it with his new book *The Cradle of Thought*. Hobson is critical of the classical piagetian answer, because it focuses on the child as a solitary mini-scientist, testing his or her current (and usually mistaken) theory in the laboratory of the sand-pit or the playroom. Quite correctly, Hobson points out the limitations of this essentially asocial perspective on cognitive development. Instead, he argues that the origins of all thought lie in social relations.

This line of argument has a long and noble pedigree. Marx suggested that all thought was a product of social and economic relations, and the Russian psychologist Lev Vygotsky (no doubt influenced by his own post-Revolution society) proposed that learning is typically facilitated by one's peer group. But Hobson's slant on the social origins of the mind comes not from this socialist framework so much as his psychoanalytic background.

Sigmund Freud, and later John Bowlby,



argued convincingly that one's earliest attachments with 'significant others' shape the development of the mind. Usually the 'attachment figure' is an adult caregiver, but — as demonstrated in the famous case of the orphaned Jewish victims of the Nazis studied by Anna Freud — in the absence of a parent figure, attachments can be equally strong towards a peer or sibling.

There is little doubt, from the thousands of experimental studies of the effects of the quality of attachment, that such early relationships are strongly deterministic of later emotional well-being. No one today (if they ever did) now questions whether abuse and neglect are bad for your later mental health — they invariably are. Hobson also reviews the intricate experiments by such pioneer child psychologists as Colwyn Trevarthen and Daniel Stern showing the exquisite sensitivity human infants have to their caregiver's emotional states and behaviour; how the 'dance' between a mother and her infant can become derailed by events such as postnatal depression.

It is not surprising that early emotional factors predict later ones. The surprise from this line of research is that early emotional factors partly predict cognitive outcomes — IQ, for example, and school attainment measures such as literacy. Reviewing the large body of evidence leads Hobson to conclude that all the unique aspects of human thought, including our capacity to use symbols, are social in origin.

He is careful to acknowledge that genetic and neurobiological factors can prevent a child from emotional engagement with others, when discussing children with the psychiatric condition of autism, with whom normal social relations are not possible. But he also reviews studies suggesting that forms of deprivation can also lead to autism, such as the sensory deprivation of congenital blindness or the emotional deprivation of children discovered during the past few years in Romanian orphanages.

There is much to admire in this immensely readable book, and Hobson is both an outstanding scholar and passionate about his subject. His human and clinical concern for people comes through clearly in his writing, and his book will be a welcome contribution to the debate in cognitive development. I part company with him on three fundamental issues, though.

First, just because some aspects of thought (such as empathy) clearly have emotional origins doesn't mean that all human thought is social in origin. How, for example, does an autistic 'savant' who can compute all prime numbers at lightning speed do this with little if any experience of emotional intimacy? This suggests to me that some aspects of cognition have little to do with social relations.

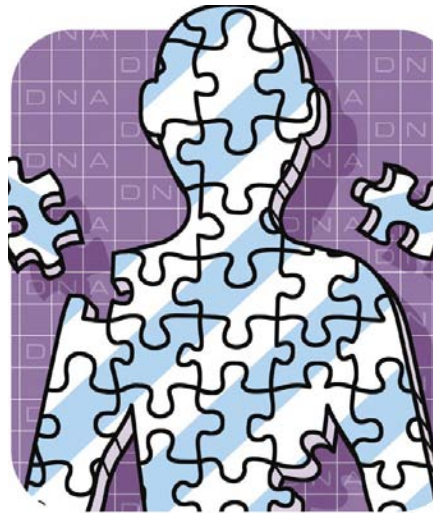
Second, although Hobson acknowledges

that genetic factors might partly determine the mind, he attacks the idea of the existence of innate, pre-programmed cognitive modules. I find this argument incoherent. Modularity theorists do not suggest that modules can function and develop without experience any more than geneticists believe that genes can function without an environment. If you accept that there are genes that can build brain structures, why not at least remain open to the possibility that genes can build mental modules?

And finally, while Hobson's groundbreaking studies of individuals with congenital blindness, or Michael Rutter's seminal studies of the Romanian orphans, have shown us that the effects of early deprivation can resemble autism, might this be no more than a surface similarity? We should be careful not to assume that just because two church bells are ringing simultaneously they are causally connected by the same rope.

Hobson's own important studies of emotion perception in autism are nicely described in this book, and in many ways were ahead of their time. There is no question that this major figure in the field of developmental psychopathology will continue to stimulate healthy debate. ■

Simon Baron-Cohen is in the Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK.



about the implications of genetic knowledge and genetic manipulation. The heat of debates on reproductive cloning and genetic testing is evidence of the anxieties of so many people. This carries a salutary lesson. If scientists misrepresent or exaggerate the power of these technologies, or are not scrupulously objective, the pursuit of knowledge is threatened. Unlike those relatively uncritical golden days of Lord Kelvin, our pronouncements will have a profound effect on public perception and the health of science, and thus on society.

One oddity about the debate on the interface between genetics and human embryology is that it has often been wrongly focused. Given the risks of producing an abnormal child — and the litigation that would ensue — I cannot believe, for example, that human cloning will be attempted in any significant way. So the crucial issue is the use of transgenic technology. We can make transgenic animals with relative ease; and we can add, modify or knock out genes in intact mammals. The key question is whether these technologies might be used in humans.

Gregory Stock of the UCLA School of Medicine bravely predicts this future for humanity. He is undeterred by the poor record of futurology, believing that it is only a matter of time before human germline modification becomes a fact. His new book, *Redesigning Humans: Our Inevitable Genetic Future*, is a distillate of opinions he has publicly expressed for some years. He writes with a clear, lucid style that lends plausibility to his views. Yet many readers will wonder whether the assertions about reproductive technology that are crucial to his argument are accurate. For example, he makes claims for the profound global impact of contraceptive technology. Without the worldwide access to birth control, Stock asserts, birth rates would not be falling. But this does not stand up to scrutiny. Falling birth rates are more to do with improved social infrastructure — better hygiene, education, decreasing infant mortality and changing social attitudes — than with his technocentric approach.

People will want genetic choice, claims Stock. He seems to believe that human nature will change so much that assisted reproduction could replace procreation on the hearth-rug. We have, he says, now accepted much of what is ethically debatable in the area of genetic choice, by using pre-implantation genetic diagnosis (PGD) in embryos. PGD will be “in the vanguard of genetic choice, at least for the next couple of decades,” Stock contends. But this is not likely to be even approximately true. Only one-fifth of embryos resulting from *in vitro* fertilization are viable; many, if not most, human embryos are frequently aneuploid or have other cellular abnormalities that are probably incompatible with development. Mosaicism is extremely common — perhaps 75% of morphologically normal human embryos have at least one or two aneuploid cells at around the eight-cell stage. Biopsy of such cells will be likely to give useless clinical results, and PGD biopsy of a normal cell in such an embryo may lead to false diagnosis.

So if *Redesigning Humans* is wide of the mark when discussing technology that has already been used for over a decade, why should futuristic comments about germline modification be any closer to the truth? Stock rightly observes that current transgenic manipulation is unpredictable, but he seems overimpressed by recent developments. It will be relatively easy to introduce auxiliary chromosomes into the germ line, and they could carry large chunks of DNA without the limitations mostly imposed by conventional gene vectors. Genes on these chromosomes could be introduced without changing other parts of the genome and could incorporate a mechanism for terminating expression to improve safety.

Stock argues that this strategy could eventually be used to enhance “desirable” characteristics for single generations and that outdated auxiliary chromosomes could be jettisoned for newer, more up-to-date models. Work in mice, he feels, suggests that this could eventually be done without human harm. But many geneticists will feel queasy; the potential for gene imbalance is huge and the change in phenotype unpredictable.

This is an important debate, but a real moral perspective is missing in this mostly engaging book. Stock favours human genetic enhancement. He quotes James Watson: “If we can make better humans... why shouldn't we?” And he is scathing about the conservative attitude of notable scientists such as French Anderson. To many readers elsewhere, his view will seem centred on privileged North America, taking little cognisance of the appalling inequalities in his and their society which would be increased by this manipulation. We are as much the product of our environment as of our genes, and much should be done first about the

Improving on humanity?

Redesigning Humans: Our Inevitable Genetic Future

by Gregory Stock

Houghton Mifflin: 2002. 288 pp. \$24

Robert Winston

William Thomson (later Lord Kelvin) entered Glasgow University at the age of ten, achieved a first in mathematics at Cambridge, published over 600 scientific papers, and became president of the Royal Society in 1890. He was a pioneering physicist but, like many competent scientists, he was not a brilliant futurologist. Less than ten years before the Wright brothers flew he said: “I can state flatly that heavier-than-air flying machines are impossible.” And he once claimed: “X-rays will prove to be a hoax.” Given his views on creationism — “overwhelming strong proofs of intelligent and benevolent design lie around us” — one wonders what he might of made of the implications of modern molecular biology.

Our imperfect knowledge of DNA and the human genome raises more unanswered questions than any other aspect of science. People from all walks of life are nervous

poor environment in which so many humans exist. For scientists to advocate channelling resources uncritically, as this book seems to do, is to risk bringing genetic research into public disrepute.

Whatever the true mechanism, perhaps Lord Kelvin was not so misguided after all in his faith in the goodness of "benevolent design". As a Christian he might have asked: if we change the very nature of what it is to be a human, will we still have our humanity? ■

Robert Winston is at the Institute of Reproductive and Developmental Biology, Hammersmith Hospital, Imperial College, London W12 0HS, UK

Flexing our muscle power

Prime Mover: A Natural History of Muscle

by Steven Vogel

W. W. Norton: 2002. 351 pp. \$25.95, £19.95

R. McNeill Alexander

Steven Vogel tells us that his professional biases, as a biologist, "start with the belief that we just can't understand history, literature, economics, art, and so forth without taking biology into consideration". Accordingly, in his book he asks what muscle physiology may reveal about human history, prehistory and culture. He is concerned with the human body, with the tools, vehicles and weapons that are powered by our muscles rather than by motors, and with the use that people make of the muscle power of animals.

To do this effectively in a book for general readers, Vogel has to use six of his 15 chapters to describe basic muscle physiology. He explains the sliding-filament mechanism of muscle contraction, the relationship between force, speed of shortening of muscle fibres and power output, and the energy cost and (in)efficiency of muscular work. He tells us about the red muscles, such as the breast meat of pigeons, that can continue working for long periods, and the white muscles of chicken breast, which are good only for short bursts of activity. We learn how the reflexes work, helping us to control movements, and how a slender muscle can exert large forces if it is built from a very large number of short muscle fibres converging on tendons.

All this is done clearly, in the readable prose that distinguishes Vogel's books, and with an attractive historical perspective. As well as being a distinguished research scientist in the field of biomechanics, Vogel is one of the best semi-popular writers on biology. He has an outstanding ability to be simple and entertaining without being misleading.

By the end of chapter 6, Vogel has got us through the basics. He has also dealt with a few of the implications of the physiology,

which are the main topic of his book. For instance, we know about the extraordinary surgical operation of cardiomyoplasty, in which a dispensable shoulder muscle is wrapped around an ailing heart and stimulated to help it pump blood. And we have been shown how the elastic recoil of stretched tendons can save energy in running, or give a catapult-like boost to a jump.

Up to this point, general readers will have learned a lot but physiologists very little. The rest of the book is written at the same non-technical level, but will give almost everyone something to think about. Vogel discusses a wide variety of hand tools, using the physiology explained in the earlier chapters in conjunction with simple engineering mechanics. He discusses our use of tools ranging from pliers to corkscrews, which amplify the force that we can exert. We discover why wood is a good material for axe handles and why modern metal axe heads work better than primitive stone ones. Vogel attributes the success of eighteenth-century American colonists to their invention of axe heads that had the hole for the shaft closer to the cutting edge than in the traditional design at the time; new-style axes were less liable to be twisted by an ill-directed blow. The author tells us why crosscut saws have two kinds of tooth, and big gaps between groups of teeth. And screws, we are told, have right-handed threads because the relative strengths of different arm muscles enable a right-handed person to twist them more forcibly clockwise than anticlockwise. In some places, scientist readers may be frustrated by a lack of detail. For example, they may want a proper explanation of the sweet spot of a hammer or baseball bat, or more information about tests of human strength. However, there are plenty of footnotes and references to the literature.

Vogel regards the wheelbarrow as one of the all-time great inventions. It enables us to move heavier loads than we could carry; it has just three supports, so that it rests steadily on uneven ground; it is highly manoeuvrable; and it offers little resistance to forward movement.



Elsewhere, cycling and rowing are discussed, albeit in less detail than I would have liked, together with less-familiar man-powered vehicles, including the ancient Greek warship known as the trireme, and pedal-powered aircraft.

Vogel gives us a great deal of fascinating discussion of obsolete technology. He discusses the relative merits of horses and oxen as draft animals, and explains the design of harness and of the ingenious whippetree, which makes all the animals in a team pull with equal force. He compares medieval machines for propelling rocks into besieged cities, calculating their ranges and the time needed for reloading. He points out that it would have been best to use the heaviest possible rocks so as to get as much momentum as possible for a given energy input. And he tells us why it made ergonomic sense to cut huge blocks of stone for the pyramids but to make small mud bricks for the Great Wall of China.

Finally, Vogel considers muscle as meat. He compares the energy content of different meats, pointing out the huge differences resulting from their different fat contents. We learn why *filet mignon* (the psoas muscle) is particularly tender, how best to cook squid, and why (according to cannibals) human flesh tastes like pork, but sweeter.

This is a book that should be enjoyed and understood by intelligent non-scientists as much as by scientists. It offers thoughtful insight into a remarkable range of past and present human activities. Physiologists may want to skip the early chapters and may regret the lack of technical detail, but they, too, should find the book entertaining and illuminating. ■

R. McNeill Alexander is in the School of Biology, University of Leeds, Leeds LS2 9JT, UK.

The tale of the human genome

The Common Thread: A Story of Science, Politics, Ethics and the Human Genome

by John Sulston and Georgina Ferry

Bantam: 2002. 320 pp. £17.99

Sydney Brenner

Much has been written about the human genome and the project to sequence it, but this is the first book by one of the scientists who played a large part in getting it done. John Sulston, former director of the Sanger Centre at Hinxton near Cambridge, UK, where much of the work was done, has written an account of the project. He tells how he moved from organic chemistry to his decisive research on the cell lineage of the nematode *Caenorhabditis elegans*, and then

New Journals

This year, *Nature's* annual new journals review supplement will appear in the 7 November issue. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 2000 and published at least four separate issues by the end of June 2002.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.
- Frequency of publication must be at least three times a year.
- The main language must be English.
- Deadline for submissions is 15 July 2002.

For each eligible title, please send at least four different issues (the first, the most recent and any two others), together with full details of subscription rates, to Mary Purton, *Nature*, The Macmillan Building, 4 Crinan Street, London N1 9XW, UK. Tel: +44 (0)20 7843 4567. Fax: +44 (0)20 7843 4596/4763. E-mail: m.purton@nature.com

sequenced its genome with Bob Waterston, before building the large sequencing factory that would work on the human genome. The book is co-written with science journalist Georgina Ferry.

From the very start, the architects of the Human Genome Project found themselves bedevilled by two problems. The first was how to obtain the very large resources required for the sequencing and to ensure the support of the scientific community. Second, they had to deal with several issues that are extraneous to scientific and technical matters, including patenting and the use of human genetic information. Many of those who worked on the project had previous experience of these problems, which arose with the invention of methods of cloning DNA; most of these problems are still with us today.

Many scientists were opposed to the project, partly because they saw sequencing as a mindless activity, but largely because it would be Big Science and that meant Big Money. They were worried that the project would drain resources away from their research, which is why additional support had to be generated that would not compete with existing research funds. Today the project is widely accepted, but at the outset almost everybody involved had to be dragged screaming into it. It was difficult for most scientists to understand that genome sequencing would give us a new approach to genetics; few had experienced the illumination I had in 1977 on seeing how Fred Sanger's sequence of bacteriophage lambda gave us the amino-acid sequences of the proteins made by all of the genes of its 50-kilobase genome.

Sulston sees the public project as a crusade against those bent on creating a monopoly in human genome data by patenting. Chief among these is Craig Venter, who set up The Institute of Genome Research (TIGR), which was financed by Human Genome Sciences, a private company that patented the sequences emerging from TIGR's expanded

cDNA sequencing programme. The genomics 'goldrush' was created when the rights for these sequences were sold to SmithKline for \$125 million.

Other companies followed in their wake, including Incyte, which also patented sequences and kept its databases closed to all except paying customers. Sequences of human genes had been patented by many working in the public research sector, but what was new and different in the case of the human genome is that the monopoly might involve every single gene, and thus every single protein, in the human body.

In 1998 Venter announced that he was setting up Celera to sequence the human genome, and that he would do it well ahead of the public-sector project by using a total shotgun strategy. The public-sector group had already departed from their original pristine standards of a genome finished to 99.99% accuracy by contemplating in 1995 the production of a 'draft' sequence. Sulston tells how the public project's work towards the draft sequence was accelerated in response to Celera's challenge, and how Sulston was able to convince the Wellcome Trust to increase its funding.

The race was finally declared a draw on 26 June 2000, when a joint announcement was made about the draft sequences obtained by each group. This was followed early in

2001 by the publication of two papers, one by Celera in *Science*, the other by the public project in *Nature*. Sulston recounts how this exacerbated the angry relations between the groups.

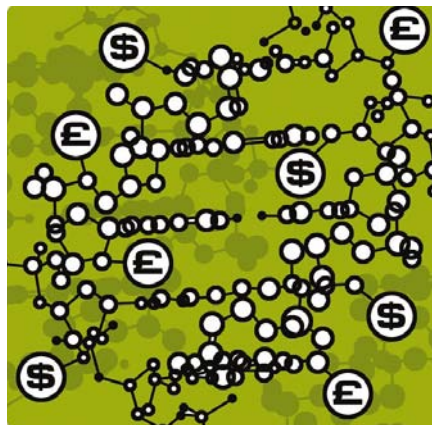
The dominant message in the book is that the human genome sequence belongs to everybody and that we should all have free access to it. Sulston formulated the Bermuda principles, which aimed to ensure this access by the daily automatic release of all sequence assemblies of more than 1 kilobase and the immediate submission of finished sequences to the public databases. The lofty aim was to make the sequence freely available for research and development, but its prior publication also rendered sequences unpatentable for lack of novelty. It did not help the ordinary scientist very much because large computing resources were required to do anything serious with the unfinished sequences. Celera had such resources, and so were able to use the public database as well as their own data.

Sequencing large genomes has nothing to do with any intellectual endeavour. The creative work was done earlier by Fred Sanger and by others who improved the technology. The rest is about two things: money and management. As the various projects developed, their demand for money increased. The nematode sequencing project quickly consumed most of the funds available for genome work at a time when money was short.

Sulston depicts Venter as a man who would sully the human genome by making people pay to use it, a devil who needed to be defeated. There is another view: that he was more Faust than Mephistopheles. He had ambitions in science but was essentially an outsider. His talents for organization were outstanding, but if he were to create a place for himself in the sequencing world he would need resources. He could not get these from the public sector, so he turned to other sources. Commercial organizations expect a return for their money, and Venter not only accepted the faustian bargain but used the deal to try to win the race. It looks like a clever business ploy to sell more machines and use the heightened interest in genome research to raise a large amount of money. Today, the genome bubble has burst and everybody has gone to look for drugs. Will the public sector carry on to finish the human genome sequence?

What I found interesting in this account is that Sulston doesn't tell us anything about the genomes he has sequenced. What did he find there that excited him? What did he learn about genes, about life, about evolution, about worlds to come? It is the play of *Hamlet* without Hamlet.

Sydney Brenner is at the Salk Institute, La Jolla, California 92037-1099, USA.



Sphere of influence

Dan Yakir

The realization that enzyme-driven molecular discrimination can be detected at the global level induces a sense of vertigo at the immensity of the scale shifts involved. Some enzymes are so central in nature that their discrimination signals are indicators of global changes. Yet, remarkably, these global-scale signals can be recreated with a few cells in a sealed test tube, or with a leaf in a bell jar.

Each element has different isotopes, most of which are stable and can be distinguished by their mass. Enzymes catalyse chemical reactions that are sensitive to differences in the dissociation energies of molecules. It is usually easier to form, or break, bonds of molecules that contain the lighter isotope of a given element, because the vibrational frequency of such bonds tends to be higher. Consequently, molecules containing lighter isotopes are preferentially incorporated into the products of incomplete reactions, whereas heavier isotopes become enriched in unreacted residues. These simple, mass-dependent traits give rise to useful signals — stable isotopes can be thought of as natural labels on all environmental processes, allowing us to see a dynamic universe that is normally hidden from view.

Take Rubisco (ribulose 1,5-bisphosphate carboxylase–oxygenase), the most common enzyme on Earth and the primary enzyme in

photosynthesis, which fixes CO_2 from the air to form sugars. Essentially, all CO_2 assimilated by photosynthesis passes through this single enzyme. Because of the mass-dependent effects described above, Rubisco discriminates against CO_2 containing ^{13}C (around 1% of carbon in nature), producing sugars that are depleted in this isotope. Because of the staggering quantities of CO_2 processed by Rubisco on the global scale, there is a marked accumulation of ^{13}C in atmospheric CO_2 .

Photosynthesis on land consumes about 350 billion tonnes of CO_2 every year (about one-seventh of all the CO_2 in the atmosphere), all of which ultimately returns to the atmosphere through respiration and combustion. On land, photosynthesis increases and decreases with the seasons, so changes in ^{13}C concentration in the atmosphere can be observed. In the ocean, CO_2 is taken up by dissolution with little discrimination. This makes atmospheric ^{13}C useful in distinguishing CO_2 uptake by photosynthesis on land and by dissolution in the ocean, a critical point in tracing the fate of CO_2 from spent fossil fuel.

Another 'global enzyme' is carbonic anhydrase, one of the fastest enzymes in nature. It catalyses the hydration of the roughly one-third of the atmospheric pool of CO_2 that is dissolved in plant and soil water each year, and in so doing imposes the ^{18}O signal of the water's origin on the CO_2 . This signal varies because of mass-dependent effects in the hydrological cycle, so carbonic anhydrase is helpful in revealing the sinks and sources of CO_2 in ecosystems and in the global biosphere, and in identifying the underlying processes.

Cytochrome oxidases, which are involved in the uptake of O_2 from the air and its reduction to form water in mitochondria during respiration, are also global enzymes. Through mass-dependent effects, these enzymes strongly discriminate against ^{18}O in O_2 (which is mostly composed of ^{16}O). Vast quantities of oxygen are processed by enzymes that preferentially convert $^{16}\text{O}_2$ to water, increasing the concentration of ^{18}O in the remaining atmospheric oxygen. Any change in the extent of this ^{18}O enrichment is an indicator of change in the 'breathing' characteristics of the biosphere (oxygen taken up in respiration is ultimately returned to the atmosphere by photosynthesis without discrimination).

Interpreting these global isotopic signals requires precise laboratory measurements. For example, ^{13}C concentrations in the sugars formed by Rubisco are always about 3% less than in the source CO_2 , and the ^{18}O content of water formed in respiration is about 2% less than in the substrate O_2 . Leaf chambers, terraria and gigantic bell jars such as Biosphere-2 are required to enhance enzyme discrimi-

Global enzymes

Stable isotopes are natural labels on all processes in the environment, allowing us to see a dynamic universe normally hidden from view.

mation sufficiently to assess their atmospheric imprints. It is these laboratory-determined isotopic discrepancies, and the ubiquity of the enzyme processes involved, that make these processes such valuable tools. Plant biologists can obtain lab results that can also be applied to large-scale transport models and general climate models that trace fluxes of CO_2 and O_2 between the atmosphere and the Earth's surface. It has been shown, for example, that land plants constitute a large sink for atmospheric CO_2 that counteracts some of the CO_2 emission associated with human activities, and that the balance between photosynthesis on land and in the ocean differs during significant climate changes, such as at the end of the most recent ice age.

Scaling enzyme reactions up to atmospheric measurements is particularly relevant at a time when biology's role in influencing atmospheric CO_2 budgets, the production of tropospheric ozone, and other climate-related phenomena is increasingly recognized. Yet the number of biologists working on global change is not increasing. This is partly because they are trained to be reductionists, dissecting complex systems into elementary parts. But global-change science is holistic, seeking to integrate components of Earth systems into regional and global patterns. Someone who is used to working at the organism or biochemical level needs to invert his or her binoculars.

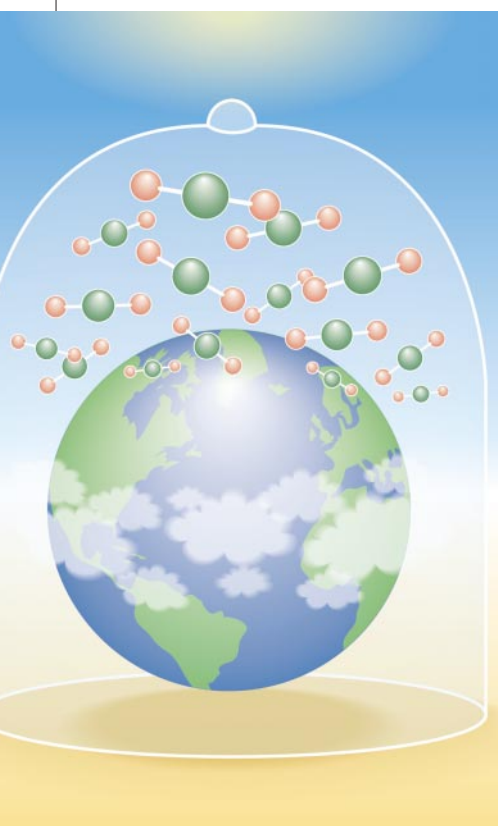
To understand how processes such as enzyme discrimination contribute to the global atmosphere, we need to include the concept of scaling and new research tools in biology training programmes. Biology can then contribute fully to the inter-, multi- and transdisciplinary research required for a complete understanding of global change. ■

Dan Yakir is in the Department of Environmental Sciences and Energy Research, Weizmann Institute of Science, Rehovot 76100, Israel.

FURTHER READING

Bigeleisen, J. *Science* **147**, 463–471 (1965).
Griffith, H. (ed.) *Stable Isotopes and the Integration of Biological, Ecological and Geochemical Processes* (Bios Scientific, Oxford, 1998).
Berry, J. A. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 411–454 (Plenum, New York, 1992).
<http://www.bio2.edu>

VICKY ASKEW



The cosmic accelerator

Felix Aharonian

The cosmic rays that permeate our Galaxy have been attributed to various sources. The discovery of high-energy γ -rays from a supernova remnant at last provides concrete evidence for one of the proposed theories.

Cosmic rays are energetic particles, mainly protons, that arrive at Earth from all directions in space. They were first detected by Victor Hess in 1912. But, despite considerable experimental data and extensive theoretical efforts over past decades, there is still no generally accepted view on where cosmic rays are produced, or how they are accelerated to such high energies. On page 823 of this issue, Enomoto *et al.*¹ (the CANGAROO collaboration) report the detection of high-energy γ -rays from the supernova remnant RX J1713.7–3946. This seemingly modest claim in fact has significant implications. If true, and provided that RX J1713.7–3946 is a typical representative of its class — the shell-type supernova remnants — it would mean that the production of cosmic rays in our Galaxy could be conclusively linked to the aftermath of supernovae, the cataclysmic explosions that mark the death of stars.

The acceleration, accumulation and effective mixing of cosmic rays, through their diffusion and convection in galactic magnetic fields, produce a 'sea' of energetic particles that pervades our Galaxy. The corresponding pressure exerted by this sea is surprisingly (but presumably not accidentally) close to the pressure of the galactic magnetic field, as well as to that of the interstellar gas. This implies that galactic cosmic rays have an important role in the dynamical balance of our Galaxy and influence interstellar chemistry through the heating and ionization of interstellar gas.

But where do the cosmic rays come from, and how are they accelerated? In 1933, Walter Baade and Fritz Zwicky² suggested a link

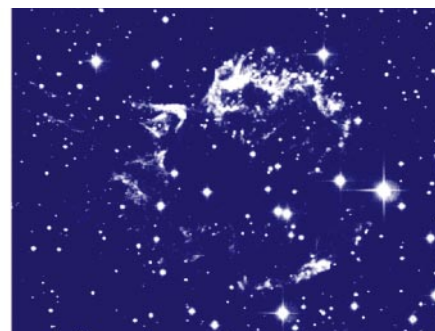
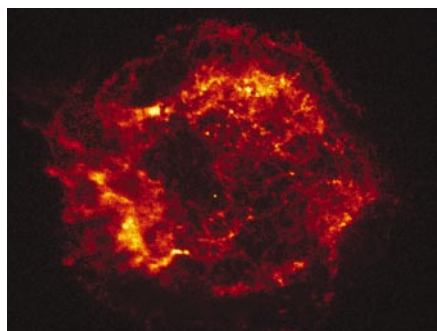


Figure 1 Seeking a source of cosmic rays. These X-ray (left) and optical (right) images show the shell-type supernova remnant Cassiopeia A — the remains of a star that disappeared 300 years ago. Astronomers suspect that cosmic rays are accelerated in the shock waves of shell-type remnants.

between cosmic rays and supernovae. The main argument is that the energy needed to maintain the galactic population of cosmic rays is estimated as at least a few per cent of the total mechanical energy released by supernova explosions in our Galaxy. This is an important, but not sufficient, argument, given that other potential sources, such as pulsars or young stars with energetic mechanical winds, might also meet this energy requirement.

The second argument in favour of supernova remnants hinges on the strong shock waves that emanate from them after the collapse of the star. According to the diffusive-shock acceleration mechanism, these shock waves could transfer 10–30% of the mechanical energy of the supernova explosion (which is around 10^{51} erg; 1 erg = 10^{-7} joule) to protons, electrons and nuclei in the surrounding interstellar gas. This mechanism can explain the energy distribution of cosmic rays that has been measured around the Earth³.

But these arguments are only theoretical; what is needed is direct empirical evidence. If protons are accelerated by the shock wave of a supernova remnant, they could interact with the surrounding interstellar gas to produce short-lived particles called π^0 mesons, which in turn would decay to produce γ -rays at very high, TeV, energies (1 TeV = 10^{12} electron volts). These γ -rays are said to be of 'hadronic' origin, after the hadron family of particles to which protons and π^0 mesons belong. If TeV-energy γ -rays from supernova remnants were detected, it would be the first experimental proof that cosmic rays are produced and accelerated in the shock waves.

In fact, γ -rays at TeV energies have been detected from two well-known shell-type supernova remnants^{4,5}, including Cassiopeia A (Fig. 1). But for these objects, it is also plausible that ultra-high-energy electrons are producing the γ -rays, by a process called inverse Compton scattering (Fig. 2): electrons accelerated by the shock can scatter photons from the cosmic microwave background (the radiation relic of the Big Bang), energizing them into high-energy γ -rays. Electrons belong to the family of particles known as leptons, so these γ -rays are said to be of 'leptonic' origin.

Meanwhile, no TeV emission has been detected from other supernova remnants (such as γ Cygni or IC 433) where protons, rather than electrons, are expected to dominate, and many consider this to be a failure of the supernova-remnant theory for the origin of galactic cosmic rays. But this conclusion is premature, given the limited sensitivity of existing detectors, as well as the large uncertainties in some key parameters in the proposed model.

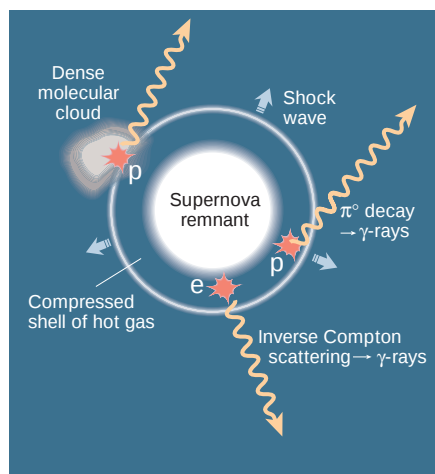


Figure 2 Getting up to speed. Shock waves from supernova remnants can produce high-energy γ -rays by two routes: shock-accelerated electrons energize photons from the cosmic microwave background to produce γ -rays by inverse Compton scattering; or accelerated protons interact with the surrounding gas to form π^0 mesons, which produce γ -rays. Enomoto *et al.*¹ show that γ -rays from the supernova remnant RX J1713.7–3946 are due to accelerated protons. This is the first strong evidence that supernova remnants are accelerators of cosmic rays (which are mainly protons). The high flux of the γ -rays detected suggests that protons are being accelerated in a high-density region, where a molecular cloud has been overtaken by the expanding shell of the supernova remnant.

Enomoto *et al.*¹ at last provide strong evidence that galactic cosmic rays do originate in supernova remnants. They have detected TeV-energy γ -rays from the supernova remnant RX J1713.7–3946 and show that the energy spectrum of the γ -ray signal matches that expected if the γ -rays are of hadronic origin, but does not agree with the spectrum expected for leptonic γ -rays (see Fig. 2 on page 824). The spectrum seen could not be produced by inverse Compton scattering — that would require a very low magnetic field (just a few microgauss) around the supernova remnant, which is unlikely to be the case. When earlier X-ray data are also taken into account, it seems that the γ -rays must have been created as a result of protons being accelerated by the supernova remnant.

The flux of TeV γ -rays reported by Enomoto *et al.* means that γ -ray production must occur in a very dense region — otherwise, the accelerated protons would have to be carrying much more energy than expected, more than 10^{50} erg, assuming that around 10% of the explosion energy is converted to cosmic-ray energy³. A nearby, massive molecular cloud — more than 10^5 times greater than the mass of our Sun — interacting with the supernova remnant⁶ would produce such a high-density environment (Fig. 1).

Remarkably, the EGRET instrument aboard the Compton Gamma Ray Observatory has detected a source of lower-energy γ -rays (3EG J1714–3857) in the vicinity of this cloud. It has been suggested⁷ that the EGRET source is indeed in the molecular cloud, and that the low-energy γ -rays are being produced through the interactions of accelerated protons with the dense cloud. In the CANGAROO data, the TeV γ -rays do seem to originate in the region of the molecular cloud and the EGRET source, but the data are of insufficient quality to pinpoint the exact location of the TeV source.

Observations with new telescopes — the European HESS and MAGIC, the Japanese–Australian CANGAROO-3 and the American VERITAS — should reveal more details about energy spectra and production sites of γ -rays in supernova remnants across a broad energy range, from 100 GeV to 10 TeV. By providing unique information about the acceleration and propagation properties of high-energy particles in several remnants, these telescopes will make a crucial contribution to our efforts to trace the origin of galactic cosmic rays. ■

Felix Aharonian is at the Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, D-69029 Heidelberg, Germany.
e-mail: felix.aharonian@mpi-hd.mpg.de

- Enomoto, R. *et al.* *Nature* **416**, 823–826 (2002).
- Baade, W. & Zwicky, F. *Phys. Rev.* **45**, 138 (1934).
- Drury, L. O'C. *et al.* *Space Sci. Rev.* **99**, 329–352 (2001).
- Tanimori, T. *et al.* (CANGAROO collaboration) *Astrophys. J.* **497**, L25–L28 (1998).
- Aharonian, F. A. *et al.* (HEGRA collaboration) *Astron. Astrophys.* **370**, 112–120 (2001).
- Slane, P. *et al.* *Astrophys. J.* **525**, 357–367 (1999).
- Butt, Y. M. *et al.* *Astrophys. J.* **562**, L167–L171 (2001).

Mammalian evolution

Upwards and onwards

Anne Weil

A newly described fossil sits on one of the lowest branches of the placental-mammal family tree. But its paws and claws suggest that, where actual vegetation was concerned, it could climb further than its contemporaries.

The fossil *Eomaia scansoria*, discovered in the Yixian Formation of Liaoning Province, China, and described by Ji *et al.* on page 816 of this issue¹, is the earliest known member of the lineage leading to placental mammals. *Eomaia* — ‘Dawn Mother’ — is exceptionally well-preserved for a 125-million-year-old. Although the fossil’s skull is squashed flat, its teeth, tiny foot bones, cartilages and even its fur are visible (Fig. 1).

This might not be surprising, considering that insects and dinosaur feathers have already been described from the Yixian Formation, and the discovery of thick fur on even a 125-million-year-old mammal is certainly not controversial. Although feathered dinosaurs surprised some, everyone always expected the earliest mammals to have hair. But most mammals that are this old are known only from a few teeth. What is so special about *Eomaia*, a dainty creature the size of a large mouse, is the overall detail of preservation, which makes it a treasure trove of information about early mammalian evolution and ecology.

Most mammals living today, including humans, nourish their young during extended gestation, through a placenta. This is in contrast to marsupial mammals and the egg-laying monotremes. *Eomaia* is the oldest representative found thus far of the lineage — Eutheria — in which placental reproduction evolved. This conclusion is unusually robust, based as it is on many dental and skeletal characteristics. Phylogenetic analysis also confirms that *Eomaia* is not just old, but one of the most primitive eutherians. Although part of the lineage leading to placental mammals, it probably reproduced more like living marsupials. Like other early mammals it has features — such as epipubic bones extending forward from its pelvis, and a narrow pelvic outlet — which suggest that a short gestation period was followed by parental nurture of the young suspended from the abdomen².

In some other respects, *Eomaia* seems to be specialized for a particular lifestyle. Ji and colleagues have studied its skeleton to make inferences about its locomotion, and they conclude that *Eomaia* was scansorial, or adapted for climbing. It is tricky to make this kind of judgement in the case of very small mammals, many of which are scansorially adapted even if they never climb vegetation or cliffs. Such animals have to scramble over roots and rocks that humans stride over, and substrates that feel flat to our big feet present



Figure 1 **Half of the fossil of *Eomaia scansoria*¹, which is preserved on two facing slabs of rock from China's Yixian Formation. The length of the animal, tip-of-the-nose to tip-of-the-tail, when uncurled, would have been about 16 cm. The dark area around the skeleton is carbonized fur. Ji and colleagues' reconstructions¹ of the fossil as a whole appear on page 817.**

them with many irregular surfaces³. So we would expect there to be little morphological distinction between a small mammal that stays on the ground and one that climbs more ambitiously. For instance, both might have long tails to help them balance, and flexible wrists and ankles that allow them to put their paws down securely at almost any angle³. But *Eomaia* does seem to have specializations indicating that it did more than navigate rough terrain and occasionally dash up a tree. Its finger and toe bones are elongated relative to its palms and soles, a feature associated with grasping twigs⁴. And on both its hands and its feet, *Eomaia*'s claws are strongly curved and laterally compressed. Living mammals with similar adaptations, for instance dormice⁵, feed and even nest in trees.

Eomaia is the most striking mammalian fossil yet to be found in the Yixian Formation. But it is not the first. The mammalian fauna is taxonomically diverse, having previously yielded the new fossil genera *Jeholodens*⁶, *Repenomamus*⁷ and *Zhangheotherium*⁸. These are representatives of other mammalian lineages that, unlike Eutheria,

have not survived to the present day. Indeed, the Yixian provides a snapshot of a time when, unlike today, the eutherians were but one of many early mammalian lineages, with no special guarantee of success or even of long-term survival. The degree of preservation of these fossils makes this picture especially valuable; although we often compare the teeth of contemporaneous animals, and draw general conclusions about diet, this fauna consists of detailed, articulated skeletal material that allows palaeontologists to visualize the niches these early mammals might have occupied. It is particularly fortuitous because the teeth of all four are consistent with carnivorous and insectivorous diets.

The four genera are morphologically — and, by inference, ecologically — quite different. For one thing, they are all of different sizes. *Repenomamus*, with a skull almost 11 cm long⁷, was a giant among Mesozoic mammals, and probably exploited different resources than its much smaller relatives, any of which it could have eaten in one bite. *Jeholodens*, similar in some respects to *Repenomamus*, but a little smaller than *Eomaia*, has relatively short fingers and toes, and claws that are too broad in cross-section to have provided a particularly good grip. *Jeholodens* has been interpreted as ground-dwelling^{1,6}. *Zhangheotherium* was larger than *Eomaia*, but fossils of it are not as complete and its feet are not well known. Its hands, however, have relatively short, straight fingers, with gently curved, broad claws. Again, it seems not to have spent as much time clambering in bushes or trees as *Eomaia* did^{1,8}. *Eomaia* may have reached food sources that were not available to other mammals or have scampered overhead to avoid predation, or it may have combined both activities.

Was climbing somehow related to the evolutionary success of the eutherians? For the moment, we can only speculate. One fossil, no matter how good it is, is scanty evidence. The next eutherian skeletons we know of were preserved roughly 50 million years after *Eomaia*, by which time eutherians had diversified to include forms that were specialized for running and hopping⁹. On the other hand, primitive members of the marsupial lineage were also climbers, and it has long been argued that the common ancestor of both groups might have been scansorial⁹ — an argument that Ji and colleagues' interpretation¹ of *Eomaia* would tend to support. Molecular and fossil evidence indicate that both lineages diversified considerably during the Cretaceous period, between 125 million and 65 million years ago. It seems at least possible that a propensity to climb gave them the upper hand. ■

Anne Weil is in the Department of Biological Anthropology and Anatomy, Duke University, 08 Bio Sci Building, Science Drive, Box 90383,

Durham, North Carolina 27708-0383, USA.

e-mail: annew@duke.edu

1. Ji, Q. *et al.* *Nature* **416**, 816–822 (2002).
2. Novacek, M. *et al.* *Nature* **389**, 483–486 (1997).
3. Jenkins, F. A. Jr & Parrington, F. R. *Phil. Trans. R. Soc. Lond. B* **273**, 387–431 (1976).
4. Lemelin, P. J. *Zool.* **247**, 165–175 (1999).
5. Kielan-Jaworowska, Z. & Gambaryan, P. P. *Fossils Strata*

36, 1–92 (1994).

6. Ji, Q., Luo, Z.-X. & Ji, S.-A. *Nature* **398**, 326–330 (1999).
7. Wang, Y.-Q., Hu, Y.-M., Meng, J. & Li, C.-K. *Science* **294**, 357–361 (2001).
8. Hu, Y.-M., Wang, Y.-Q., Luo, Z.-X. & Li, C.-K. *Nature* **390**, 137–142 (1997).
9. Kielan-Jaworowska, Z., Bown, T. M. & Lillegraven, J. A. in *Mesozoic Mammals* (eds Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A.) 221–258 (Univ. California Press, 1979).

Semiconductors

An eye for impurity

Paul S. Peercy

As electronic devices become smaller, so the challenge of maintaining their electrical properties grows. Identifying the positions of introduced impurities in a semiconductor crystal is a major first step.

The electrical properties of a semiconductor can be controlled by adding certain foreign atoms to the material. This technique is called impurity doping. The dopant atoms are carefully placed on lattice sites (the locations occupied by atoms in the crystal) in place of the atoms of the semiconductor. Each dopant impurity atom contributes one free electron or hole to the conduction or valence band, respectively; this holds over a large range of dopant concentrations. Impurity doping is becoming an increasingly difficult problem for the semiconductor industry, for the key linear relationship between dopant concentration and free-electron concentration breaks down when the concentration of dopant atoms is high. But as integrated circuits become smaller, the charge-carrier concentration must be increased to maintain the required device performance.

In an effort to understand, at the atomic level, why the free-carrier concentration saturates in this way, Voyles and colleagues¹ (page 826 of this issue) have succeeded in imaging single impurity atoms bonded in a silicon crystal. Silicon is the mainstay of the microelectronics industry. Dopants such as phosphorus or antimony contribute negative charge carriers (electrons) to the conduction band, making a region of the crystal 'n-type'; dopants such as boron or aluminium contribute positive charge carriers (holes) to the valence band, making a region of the semiconductor crystal 'p-type'.

The next few years will see the need for ever higher dopant concentrations as device dimensions continue to be scaled down. In fact, the required concentrations exceed the maximum allowed if the impurity atoms are to be thermodynamically stable in the host material². Technical tricks have already been developed that create these supersaturated impurity concentrations with remarkable control and reproducibility — for example, ultra-rapid quenching to 'freeze' the solution of impurity atoms in place³.

But researchers have found⁴ that, as dopant-atom concentrations approach the

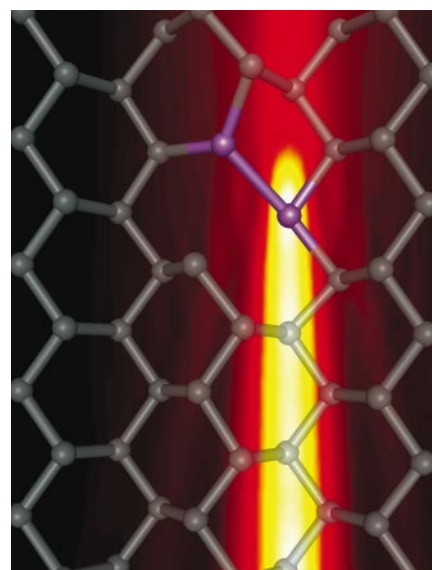


Figure 1 The electron probe of a scanning transmission electron microscope propagates along a column of silicon atoms in the semiconductor lattice. The colour scale (white to red) indicates the decreasing intensity of scattered electrons; a ball-and-stick representation of the silicon lattice, including antimony dopant atoms (purple), is superimposed. The electron-scattering intensity is highest around an antimony atom, because it has a higher atomic number than silicon — Voyles *et al.*¹ have exploited this effect to image individual dopant atoms in the lattice. Understanding the behaviour of dopant atoms at high concentrations is vital for the semiconductor-based electronics industry.

high levels required for smaller devices — around 10^{21} dopants per cm^3 — the electron concentration in the conduction band does not exceed $6\text{--}7 \times 10^{20}$ electrons per cm^3 . This means that, at high dopant concentration, a large fraction of the dopant atoms are not electrically active, and we need to understand why.

Voyles *et al.*¹ set out to locate the dopant atoms in silicon crystals that had been doped with the metal antimony. Locating individual

antimony atoms among all the silicon atoms in the crystal is like searching for the proverbial needle in a haystack. To 'see' the antimony dopant atoms, the researchers adapted a powerful technique — scanning transmission electron microscopy⁵ — to image almost all the antimony atoms in the silicon crystal, even individual antimony atoms.

The first critical step was to prepare thin crystals — only 9 or 11 atoms thick — lacking any surface or near-surface disorder that could scatter the incoming electron beam of the scanning transmission electron microscope and prevent it from being guided through the channels formed by the rows of atoms in the lattice. The cross-section (or probability) for electron scattering increases with the atomic number, Z , as $Z^{1.7}$, so an antimony atom, with $Z=51$, scatters electrons nine times more strongly than a host silicon atom with $Z=14$. By detecting the intensity of the scattered electrons, Voyles *et al.*¹ could readily distinguish rows in the atomic lattice that contained a single antimony atom from rows containing only silicon atoms (Fig. 1).

From these elegant measurements, Voyles *et al.* conclude that the electrically inactive atoms are grouped in clusters containing two antimony atoms. But their measurements also revealed that the basis on which impurity dopant atoms remain isolated or form small

clusters is random. This suggests that overcoming carrier saturation at very high dopant concentrations will be a major challenge.

Seeing is frequently the first step towards understanding. This first, unambiguous observation of single atoms bonded inside a bulk solid environment has wide-ranging implications for the analysis of single atoms and of clusters of two, three or four atoms — and for our efforts to understand the structure of impurities and alloy constituents in crystalline solids. Voyles *et al.*'s results¹ are important in understanding the distribution of impurities in silicon at an atomic level;

they will also be important in increasing our understanding of a wide range of complex materials. ■

Paul S. Peercy is at the College of Engineering, University of Wisconsin, 1415 Engineering Drive, Madison, Wisconsin 53706-1691, USA.
e-mail: peercy@engr.wisc.edu

1. Voyles, P. M., Muller, D. A., Grazul, J. L., Citrin, P. H. & Gossman, H.-J. L. *Nature* **416**, 826–829 (2002).
2. 2001 International Technology Roadmap for Semiconductors (Semiconductor Industry Association, 2001); www.semichips.org/pre_technology.cfm
3. Gossman, H.-J., Rafferty, C. S. & Keys, P. *Mat. Res. Soc. Symp.* **610**, B1.2.1–B1.2.10 (2000).
4. Packan, P. A. *Science* **285**, 2079–2080 (1999).
5. Pennycook, S. J. & Narayan, J. *Appl. Phys. Lett.* **45**, 385–387 (1984).

Plant biology

On guard

Pierre J. G. M. de Wit

Microorganisms that infect plants must suppress their hosts' defence mechanisms before they take up residence. But some plants use molecular guards to sense when they are being manipulated by pathogens.

As a rich source of sugars and amino acids, plants attract a variety of intruders, from viruses, bacteria and fungi to insects. To protect themselves, plants have in their armoury of passive defence mecha-

nisms such as strengthened cell walls and antimicrobial compounds, as well as active healing responses. Only a few microbes can breach these 'basal' defences, and are then fought by the plant's innate immune system.

Agriculture

Champagne surprise

Pinot noir is an important grape variety that is used for making wine, including champagne. It is derived from an ancient strain that has been cultivated for maybe 2,000 years. Cultivated grapevines (*Vitis vinifera*) are usually reproduced from cuttings, so all individuals are genetically identical. But sometimes mutations arise, and long ago this resulted in the generation of another champagne grape variety, Pinot Meunier (pictured here), from Pinot noir. Pinot Meunier plants are genetically indistinguishable from Pinot noir in most cells, but their outer layer, the 'L1' epidermal cell layer, is different — meaning that Pinot Meunier has a furry surface on its leaves whereas Pinot noir does not. Elsewhere in this issue (*Nature* **416**, 847–850; 2002), Paul K. Boss and Mark R. Thomas describe the precise mutation that causes this difference. Surprisingly, it is the grapevine equivalent of the 'dwarfing' mutations used to increase wheat yields during the green revolution.

Boss and Thomas started by producing grapevines that carried the mutation in all their cells, not just

their skin: using tissue culture, the authors regenerated whole plants from Pinot Meunier L1 cells. As well as having hairy leaves, these plants were semi-dwarfed — they were shorter and stockier than usual. This provided a clue about which gene might be mutated. Sure enough, it turned out to be the grapevine equivalent of the gene that, when mutated, causes dwarfing in wheat. Gene sequencing showed that the gene had a similar mutation in the L1-cell plants and in dwarfed wheat.

What does the gene do? It was first identified and cloned from the thale cress *Arabidopsis thaliana*, and named *GA INSENSITIVE* because of the effect on the plant of mutating it. It is a regulatory gene that normally keeps a brake on plant growth; the brake is released by gibberellic acid (GA). Thus the plant can regulate its growth by controlling the production and location of this hormone. Some mutations in the gene disrupt the encoded protein so that gibberellic acid no longer releases the brake on growth. This means that the brake is permanently on, and the plant is



smaller. Of course this only occurs if all cells are mutant; if the mutation is limited to the epidermis the only change seen is increased hairiness — gibberellic acid presumably also suppresses hair growth.

There was one more surprise from Boss and Thomas's study: in the L1-cell plants, tendrils were replaced by flowering stems. Normally, a new shoot produces several bunches of flowers (and thus grapes) opposite the first few leaves, and tendrils opposite leaves that form later. The tendrils anchor the vine as it grows in search of light. But in the semi-dwarfed mutant, flowering stems continued to form in place of tendrils. Presumably, the explanation is again that the plants cannot respond correctly to gibberellic acid. Normally,

this hormone may ensure that the later-arising structures become tendrils rather than flowering stems. Gibberellic acid was known to influence flowering in other plant species, but this function is apparently new. Interestingly, tendrils in some plants such as peas have a different origin — they are modified leaflets rather than flowering stems. New knowledge about hormone-response genes may allow us to fine-tune both the vegetative architecture of grapevines, and how many bunches of grapes they produce.

David R. Smyth is in the School of Biological Sciences, Clayton Campus, Monash University, Victoria 3800, Australia; e-mail: david.smyth@sci.monash.edu.au

NATURE

100 YEARS AGO

Our present state of civilisation has of necessity resulted in an annual increase in the amount of capital borrowed by man from the store of energy accumulated by our earth in bygone times, and the diversion of this capital to uses for which the world's annual income of solar energy was formerly deemed adequate. An instance of this tendency is afforded by the experiments of Dr. Selim Lemström, of Helsingfors, on the uses of electricity in stimulating the growth of cereals, vegetables and other plants... The investigation seems to have been suggested in the first instance by an attempt to connect the luxuriant growth of plants in high latitudes with the influence of electric currents associated with the Aurora Borealis. The experiments showed that for plants growing on arable land of medium quality an increase of 45 per cent. in the crops is obtainable; but the better the field is ploughed and cared for the greater will be the increase. On poor soil the effect is trifling. Certain plants, such as peas, cabbages and turnips, only lend themselves to electrical treatment after being watered. It is, however, injurious to most, if not all, plants to submit them to the influence of electricity in hot sunshine... A further suggestion is that we have here an explanation of the needle-shaped leaves of coniferous plants which are well adapted to facilitate the passage of electricity, or in common parlance, "attract electricity."

From *Nature* 24 April 1902.

50 YEARS AGO

Crystalline penicillin (sodium salt) was added at the rate of 12.6 gm. per ton to a standard turkey-rearing mash... The experiments briefly summarized show that the addition of small amounts of penicillin to the diet of growing turkeys not only produced a significant increase in growth-rate but also that it appeared to have a profound effect in stimulating metabolism, tiding chicks over a critical period of susceptibility to chilling and other non-specific causes of mortality during the first few weeks of life. Rate of growth and mortality during rearing are factors of primary importance to the turkey industry and largely determine the profitability of the undertaking... Its administration to birds destined for the table within the shortest possible time would appear to be strongly indicated.

From *Nature* 26 April 1952.

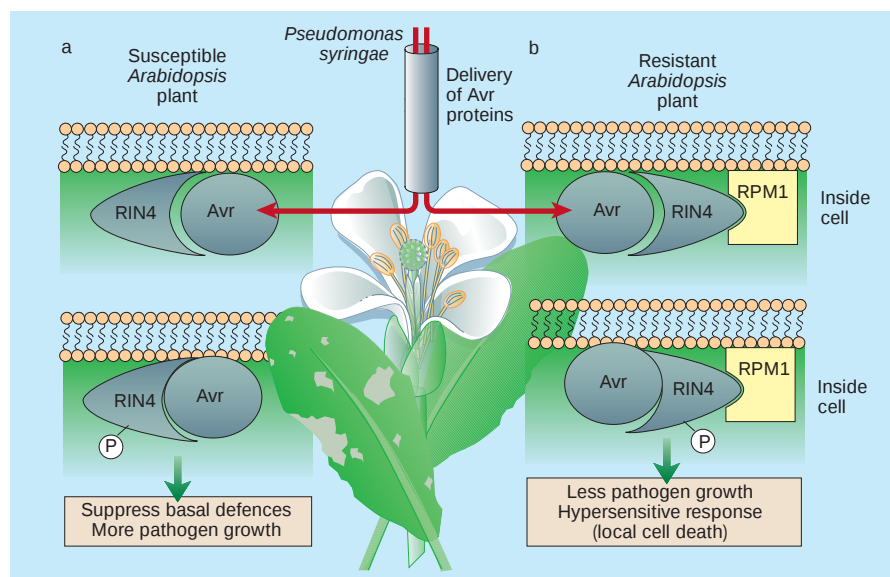


Figure 1 At the crossroads of disease resistance and susceptibility. Mackey *et al.*³ studied the *Arabidopsis thaliana* resistance (R) protein RPM1, and the *Pseudomonas syringae* avirulence (Avr) proteins AvrRpm1 and AvrB. (*P. syringae* infects *A. thaliana* leaves.) The authors found that the two bacterial proteins can bind to the plant protein RIN4, tag it with phosphate groups (circled 'P') and upregulate its concentration and its activity as a negative regulator of plant basal defence mechanisms. a, In susceptible plants, this downgrading of plant defences results in the spread of the bacterium. b, Resistant plants make use of the RPM1 protein to sense these manipulations of RIN4 by the bacterial Avr proteins, activating the hypersensitive response which prevents bacterial spread. Figure modified from ref. 3.

The innate immune response is well described genetically by what is known as the gene-for-gene model¹, because it requires a pathogen protein encoded by an 'avirulence' (Avr) gene to be recognized by a plant protein encoded by a resistance (R) gene. This activates an array of defence mechanisms, including the hypersensitive response², in which a few plant cells at the site of infection die, thereby limiting the spread of disease.

Although many Avr and R proteins have been identified, we have still much to learn about what they do. For example, do Avr proteins bind directly to R proteins? And what are the main functions of Avr proteins? They are surely not there just to enable the plant to detect the intruder. Writing in *Cell*, Mackey *et al.*³ have proposed some answers. The authors studied the R protein RPM1 from the thale cress *Arabidopsis thaliana* and two Avr proteins, AvrB and AvrRpm1, from the bacterium *Pseudomonas syringae*. They show that RPM1 is a molecular guard that prevents the bacterial proteins from taking advantage of another plant protein, RIN4, that downgrades plant basal defences. The results provide concrete support for the 'guard hypothesis' of plant defence (reviewed in refs 2,4).

Some predictions about the molecular connections between Avr and R proteins have emerged from studies of their structures and amino-acid sequences. For the most part, pathogen Avr proteins are structurally unrelated to each other. But most plant R proteins do have similar structures and contain sequences with a high proportion of leucine amino acids.

These 'leucine-rich repeats' (LRRs) are thought to be involved in protein-protein

interactions⁵, and so are expected to allow R proteins to lock directly onto Avr proteins, although (despite much searching) so far this has been shown only for the rice Pi-ta protein and the AvrPita protein from the rice-blast fungus *Magnaporthe grisea*⁶. LRRs may also specify interactions with other plant proteins (although the specificity of R proteins is not determined solely by their LRRs⁷). Some R proteins have LRRs in their extracellular regions; here, the LRRs are expected to be involved in recognizing pathogens outside plant cells⁸. By contrast, the *A. thaliana* RPM1 protein resides wholly within plant cells, associated with the plasma membrane, where it was thought to interact directly with Avr proteins like *P. syringae* AvrRpm1 and AvrB⁹. These Avr proteins are injected into leaf cells by the tubular bacterial type III secretion machinery¹⁰.

But it seems that things are not so simple. Mackey *et al.*³ find that the *A. thaliana* RPM1 and *P. syringae* AvrRpm1 and AvrB proteins interact not directly but indirectly, through the newly discovered plant protein RIN4 (for 'RPM1-interacting protein-4'). The authors first identified RIN4 because of its interaction with AvrB, and showed that it also interacts with RPM1 and AvrRpm1. They then discovered that RIN4 suppresses the expression of plant genes needed for basal defence; moreover, both AvrRpm1 and AvrB induce the covalent addition of phosphate groups to RIN4, possibly enhancing its ability to negatively regulate plant defence. Conversely, however, RIN4 is essential for RPM1 to activate the hypersensitive response to the two *P. syringae* proteins, thereby inhibiting bacterial growth — this response does not occur in

plants with low levels of RIN4. Reductions in RIN4 levels also cause reductions in RPM1 levels and (for as yet unknown reasons) resistance to both *P. syringae* and, intriguingly, the unrelated fungus-like *Peronospora parasitica*.

All of which suggests that RIN4 sits at a crossroads between susceptibility and disease resistance, and that RPM1 guards *A. thaliana* against pathogens that use AvrRpm1 and AvrB to manipulate RIN4 activity³ (Fig. 1). So, when susceptible plants are infected by *P. syringae*, the Avr proteins interact with RIN4, induce its phosphorylation, and increase its concentration, thereby inhibiting basal defences and leading to susceptibility. But in plants that are resistant to *P. syringae*, these manipulations are somehow sensed by RPM1, which launches a local cell-death programme that leads to resistance.

So Mackey *et al.*³ have shown how one R protein and two Avr proteins work at the molecular and cellular levels, causing either disease or the hypersensitive response according to the balance of power between the proteins. In so doing, the authors have answered the questions (at least for this set of proteins) of how an R protein can sense the presence of its cognate Avr proteins — through their manipulation of RIN4 — and what the Avr proteins do. It is known that some fungal and bacterial Avr proteins function in susceptible plants as 'virulence factors', thought to be required for maximum virulence of the pathogen¹¹. Mackey *et al.*³ have revealed that Avr proteins can do this by increasing the activity of a plant defence inhibitor.

This study should give a boost to those studying the molecular interactions between plants and microbes. It is likely that the direct interaction shown for the rice and rice-blast fungus proteins⁶ is the exception, not the rule; mechanisms like that described by Mackey *et al.*³ may be more common. Many labs are now hunting for virulence-related targets of Avr proteins, akin to RIN4, in other model gene-for-gene systems. It will be interesting to see whether different pathogens use the same targets. For instance, if the *P. parasitica* Avr proteins (which have not yet been identified) also interact with RIN4, that might explain why plants that are susceptible to *P. syringae* are also susceptible to *P. parasitica*. Another question is whether different Avr proteins from the same pathogen are recognized by the same plant protein, as are AvrB and AvrRpm1. I anticipate that most pathogens have a set of Avr proteins that work together to afford full virulence.

Finally, as mentioned above, some Avr proteins — such as those from the tomato pathogen *Cladosporium fulvum*⁸ — are detected by LRR motifs on the outside of plant cells, rather than inside. Might a similar guard mechanism protect against pathogens like these, too? Support for this idea comes from the finding that the hypersensitive response of tomato to one *C. fulvum*

Avr protein (Avr2)¹² requires not only the cognate R protein (Cf-2) but also a further tomato protein (Rcr3), found outside cells, which might be a virulence-related target¹³. Perhaps the R protein protects this target from the pathogen protein. Molecular guards may be widely used to prevent plant proteins from being subverted for pathogenic purposes⁴. ■

Pierre J. G. M. de Wit is in the Laboratory of Phytopathology, Wageningen University, Binnenhaven 5, 6709 PD Wageningen, The Netherlands.

e-mail: pierre.dewit@fyto.dpw.wag-ur.nl

Relativity

Testing times in space

Steve K. Lamoreaux

We take for granted that physical 'constants', such as the speed of light, are fixed values. But they might not be, and experiments in space may allow us to investigate this possibility.

A basic assumption of Einstein's theory of relativity is that the fundamental physical laws and parameters do not depend on the position, orientation or uniform velocity of the laboratory in which they are measured — a property generally known as Lorentz invariance. Relativity has been tested, implicitly and explicitly, in countless experiments; as yet, no failure of the theory has been observed. But most explicit tests have been confined to laboratories on Earth. In *Physical Review Letters*, a theoretical analysis by Bluhm *et al.*¹ shows that experiments in space — some already planned for the International Space Station — could offer better sensitivity, as well as extending the range of tests that could be performed.

Some of the most exacting tests of relativity have involved atomic clocks. These 'tick' by electrons moving between energy levels emitting a photon with a certain frequency. The tests compare the tick rates of two different atomic clocks as a function of their orientation and velocity through space. The idea is that, if the two clocks are based on different types of energy-level transitions, any failure of Lorentz invariance would show up as a relative shift in the two frequencies of the clocks, because the physical 'constants' governing the ticks of the clocks would not actually be constant, but would change with the clocks' orientation and velocity.

In Earth-based experiments, the orientation and velocity of the clocks are determined by the Earth's rotation and revolution about the Sun, and by the motion of the Solar System relative to the Universe as a whole. Typically, the differential clock frequencies are measured as a function of time. Given the Earth coordinates of the clocks, together with the time and date, the time-dependent orientation of the clock can be determined. Then,

1. Flor, H. H. *Annu. Rev. Phytopathol.* 9, 275–296 (1971).
2. Dangl, J. L. & Jones, J. D. G. *Nature* 411, 826–833 (2001).
3. Mackey, D. *et al. Cell* 108, 743–754 (2002).
4. Van der Hoorn, R. A. L., Joosten, M. H. A. J. & de Wit, P. J. G. M. *Trends Plant Sci.* 7, 67–71 (2002).
5. Ellis, J., Dodds, P. & Pryor, T. *Trends Plant Sci.* 5, 373–379 (2000).
6. Jia, Y. *et al. EMBO J.* 19, 4004–4014 (2000).
7. Luck, J. E. *et al. Plant Cell* 12, 1367–1377 (2000).
8. Joosten, M. H. A. J. & de Wit, P. J. G. M. *Annu. Rev. Phytopathol.* 37, 335–367 (1999).
9. Nimchuk, Z. *et al. Cell* 101, 353–363 (2000).
10. Hueck, C. J. *Microbiol. Mol. Biol. Rev.* 62, 379–433 (1998).
11. Leach, J. E. *et al. Annu. Rev. Phytopathol.* 39, 187–224 (2001).
12. Luderer, R. *et al. Mol. Microbiol.* (in the press).
13. Dixon, M. S. *et al. Proc. Natl Acad. Sci. USA* 97, 8807–8814 (2000).

any violation of relativity can be correlated with some aspect of orientation or velocity.

An example of a clock comparison experiment is to test whether the speed of light, c , is a universal constant — is there a limiting speed for matter, c_m , that is different from the speed of light? For instance, my colleagues and I have sought² a difference in the values of these numbers by comparing the behaviour of two atomic nuclei, ¹⁹⁹Hg and ²⁰¹Hg. In an applied magnetic field, the magnetic-moment vector of a nucleus precesses about the field direction at a particular frequency (this process is also the basis of magnetic resonance imaging). The ¹⁹⁹Hg nucleus is spherical and so its orientation in space (usually defined relative to distant, 'fixed' stars) does not affect the precession frequency. But the ²⁰¹Hg nucleus is egg-shaped — its lack of spherical symmetry means that the angle between its velocity vector and its magnetic moment becomes important; and if c_m is not identically equal to c , a shift in the precession frequency of ²⁰¹Hg compared to that of ¹⁹⁹Hg appears. In this Earth-bound experiment, no difference between c_m and c was detected, implying that if such a difference exists, then $1 - c_m/c < 10^{-22}$ — a rather astoundingly accurate limit.

This experiment is prototypical of many of the experiments described by Bluhm *et al.*¹, and we can question whether there would be an advantage to performing it in space. Among the advantages that Bluhm *et al.* specifically address is the ability to change the orientation and velocity of space-borne clocks to arbitrary directions; orientation changes could also be made more rapidly than the once-per-day change for an Earth-based experiment (which would avoid problems due to slow drift in the clock frequencies). In fact, in the specific

case of the Hg-nuclei experiment, there would only be about a factor of two to be gained if it were performed in space. But for other experiments there could be greater advantage; moreover, some tests that are impossible on Earth could be done in space.

A parameter that could be significantly varied in a space-borne experiment is the gravitational potential that the clocks are subjected to. On the basis of astronomical observations it has been claimed³ that the fine-structure constant α (which characterizes the strength of the electromagnetic interaction between photons and charged particles such as electrons) is time-dependent; and a possible dependence of α on the gravitational force has been suggested⁴ in the context of string theory — an abstract mathematical theory in which elementary particles are modelled as standing waves in a closed string loop, and which is generally beyond the reach of laboratory experiments. Whether α is dependent on gravity could be one of the few testable predictions of string theory.

A NASA team has proposed⁵ a measurement of the effect of the gravitational force on α . The project, called the SpaceTime mission, would send a spacecraft carrying three atomic clocks (Hg^+ , Cd^+ and Y^+ ions stored in three separate radiofrequency traps) to within four solar radii of the Sun. The third clock can be thought of as an on-board observer that allows the principal data analysis to be done in real-time on the spacecraft. As the craft approaches the Sun, it will experience an enormous gravitational potential for a few hours. The expected sensitivity to any change in α is around one part in 10^{16} — more than sufficient to test the string-theory prediction of one part in 10^{10} .

Could Earth-based laboratory experiments ever match this precision? String theory also predicts an annual modulation $\Delta\alpha/\alpha \approx 2 \times 10^{-14} \text{ yr}^{-1}$ because the Earth's orbit around the Sun is not circular. The most accurate laboratory tests^{6,7} have achieved $\Delta\alpha/\alpha < 4 \times 10^{-14} \text{ yr}^{-1}$, so the predicted effect is still unobservable. But there is a proposal⁸ to measure $\Delta\alpha/\alpha$ with sensitivity better than 10^{-15} yr^{-1} , which is at the level of the astronomically observed variation. This level of sensitivity would provide an Earth-based laboratory test of string theory — but still with 10,000 times less sensitivity than SpaceTime.

As Bluhm *et al.*¹ point out, space-based experiments have greater potential to discover new types of gravitational effects. For example, the SpaceTime clock-comparison experiment will be able to probe new gravitational physics far beyond what can be done in a terrestrial laboratory.

Steve K. Lamoreaux is in the Physics Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

e-mail: lamore@ziffr.lanl.gov

1. Bluhm, R., Kostecký, V. A., Lane, C. D. & Russell, N. *Phys. Rev. Lett.* **88**, 090801 (2002).

2. Lamoreaux, S. K., Jacobs, J. P., Heckel, B. R., Raab, F. J. & Fortson, E. N. *Phys. Rev. Lett.* **57**, 3125–3128 (1986).
3. Webb, J. K. *et al.* *Phys. Rev. Lett.* **87**, 091301 (2001).
4. Sandvik, H. B., Barrow, J. D. & Magueijo, J. *Phys. Rev. Lett.* **88**, 031302 (2002).
5. Maleki, L. & Prestage, J. in *Testing Relativistic Gravity in Space: Gyroscopes, Clocks, Interferometers* (eds Lämmerzahl, C.,

- Everitt, C. W. F. & Hehl, F. W.) 369–380 (Springer, Berlin, 2001).
6. Prestage, J. D., Tjoelker, R. L. & Maleki, L. *Phys. Rev. Lett.* **74**, 3511–3514 (1995).
7. Sortais, Y. *et al.* *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* **47**, 1093–1097 (2000).
8. Torgerson, J. R. Preprint physics/0012054 (2001); <http://xxx.lanl.gov>

Developmental biology

An arresting activity

Nicholas S. Duesbery and George F. Vande Woude

Vertebrate eggs pause at a crucial stage in their development, starting again only after being fertilized by sperm. Another component of the activity that ensures this arrest has been identified.

The process of vertebrate egg development consists of oocyte growth coupled with specific pauses at various stages of maturation. The final pause is known as the metaphase II arrest, from which eggs are released only after being fertilized by sperm. The cellular activity that ensures this arrest has been known for 30 years by the alias 'cytostatic factor' (CSF), ever since Masui and Markert¹ showed that egg development is arrested by CSF before fertilization. Subsequently it has been established that a signalling pathway involving the Mos protein is a key component of CSF activity. However, inactivation of CSF and the Mos pathway occurs after fertilization triggers the resumption of development, suggesting that as-yet-uncharacterized elements of CSF are required to maintain the developmental arrest. On page 850 of this issue, Reimann and Jackson² define a new, Mos-independent component of CSF, which regulates the degradation of proteins required for developmental arrest in eggs. Their work provides new insight into the regulation of CSF, and raises questions about the contribution of the Mos pathway to CSF activity.

There are two processes by which cells normally divide: mitosis and meiosis. During mitosis, cells replicate their chromosomes (producing a '4n' DNA content) and segregate them equally to each of two daughter cells (which are therefore '2n'). By contrast, during meiosis of female germ cells (oocytes), the replicated (4n) chromosomes undergo two successive 'reductive' divisions — separating first a set of chromosomes (2n) in the first polar body and second a set of replicated chromosomes after fertilization — to generate a haploid set (1n) in the egg. By contrast, male meiosis (spermatogenesis) proceeds through the first and second meiotic divisions to produce four (1n) germ cells (sperm). Progression through female meiosis (oogenesis) to produce one unfertilized egg occurs in a stepwise fashion in response to defined extracellular cues. In vertebrates, female germ cells pause just before the first meiotic division and undergo

an extended period of growth, which is necessary to accumulate resources for early embryonic stages of development. Having accomplished this, oocytes become competent to respond to hormonal signals and resume meiosis.

Resumption of meiosis is accompanied by the activation of maturation-promoting factor (MPF; Fig. 1) — a complex consisting of a regulatory protein, cyclin B2, and a catalytic protein, Cdc2. The activity of MPF is controlled by the association of these proteins with each other, and by two competing enzymes, the Myt1 kinase and Cdc25, that either add or remove phosphate groups from Cdc2. Active MPF is required for dissolution of the envelope surrounding the nucleus and formation of the 'spindle' apparatus that will segregate homologous chromosomes.

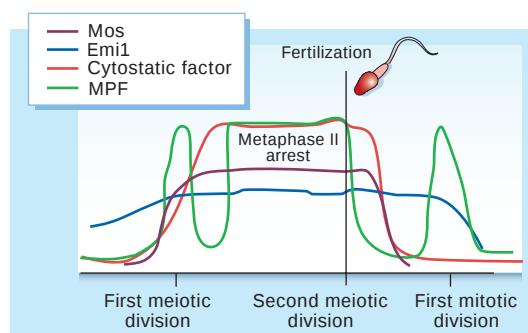


Figure 1 The ups and downs of protein activity during meiosis in vertebrate eggs. The activity of maturation-promoting factor (MPF) is carefully controlled to ensure that oocytes carry out the first division of meiosis (as MPF levels first peak), and then arrest before fertilization triggers the second division (when MPF levels are again high). Cytostatic factor (CSF) is an activity that maintains high levels of MPF in arrested oocytes. The molecules that make up CSF have been a mystery, although a role has been proposed for the Mos protein³. Reimann and Jackson² show that the Emi1 protein can block the destruction of a component of MPF (cyclin B2) and so keep MPF levels high during the arrest. Emi1 may work with Mos to induce and maintain this arrest.

Chromosome segregation is then triggered by the anaphase-promoting complex (APC), which mediates the degradation of cyclin B2 and, hence, inactivation of MPF.

Entry into the second meiotic division occurs as MPF is reactivated, when newly synthesized cyclin B2 associates with Cdc2. But now the degradation of cyclin B2 is blocked so eggs arrest division before they separate their replicated chromosomes, at a stage called metaphase II. This arrest may be a safeguard against parthenogenesis—embryonic development without fertilization—as its release is triggered by fertilization.

Our first insight into how oocytes arrest in metaphase II came from Masui and Markert¹, who showed that cytoplasm from a metaphase-II-arrested oocyte could induce a similar arrest when injected into a mitotically dividing embryonic cell (blastomere). They concluded that this arrest could be attributed to a specific activity, dubbed CSF, in the cytoplasm of the arrested oocyte. They found that CSF arose in the cytoplasm following the appearance of MPF, roughly when the first meiotic division began; and that CSF levels increased between the first and second meiotic divisions and declined after fertilization, shortly after the inactivation of MPF¹.

But attempts to biochemically purify CSF were frustrated by its instability, and the study of the molecular properties of CSF languished for almost 20 years, until Sagata *et al.*³ showed that the Mos protein has many characteristics in common with CSF. Mos was synthesized prior to entry into the first meiotic division and degraded after fertilization; injection of Mos into a mitotic blastomere caused metaphase-II-like arrest; and extracts of metaphase-II-arrested eggs that had been depleted of Mos lacked CSF activity³. Thus, Mos is a component of CSF.

Mos is a protein kinase that activates the mitogen-activated protein kinase (MAPK) pathway. Like Mos, each component of this pathway, including MAPK and p90^{rk}, behaves like CSF when injected into blastomeres^{4–6}. Most recently⁷, p90^{rk} has been shown to phosphorylate and activate the APC inhibitor Bub1. Thus, synthesis of Mos and activation of the MAPK pathway might culminate in activation of Bub1 and inhibition of APC activity, which consequently cannot trigger cyclin B2 degradation and release from metaphase II arrest.

However, although the Mos pathway is a necessary component of CSF, by itself it is insufficient to maintain metaphase II arrest; after fertilization, cyclin B2 is degraded before inactivation of the Mos pathway. Indeed, it does not even appear to be necessary to maintain metaphase II arrest, as Reimann and Jackson² now show that blocking this pathway with a MAPK inhibitor does not cause cyclin B2 degradation in CSF-containing extracts. These authors instead propose that the Emi1 protein, identified

previously as an APC inhibitor, is needed to maintain metaphase II arrest.

Reimann and Jackson found that adding extra Emi1 to the CSF-containing extracts prevented the destruction of cyclin B2 that is triggered by the addition of calcium—a process that mimics fertilization. By contrast, depletion of Emi1 led to cyclin B2 degradation. Furthermore, as shown previously⁸, adding Emi1 to mitotic blastomeres arrests their division. Their results suggest that Emi1 arrests the second meiotic division by preventing the destruction of cyclin B2, in a MAPK-independent way, and that fertilization would lead to the inhibition of Emi1 and activation of the APC, releasing the egg from metaphase II arrest.

What, then, is the role of the Mos pathway in CSF activity? The observation that mouse oocytes lacking the *mos* gene fail to arrest at metaphase II and spontaneously complete the second meiotic division^{9,10} indicates that Mos is both necessary and sufficient for entry into metaphase II arrest, especially given that naturally occurring (endogenous) Emi1 fails to cause this arrest in the absence of Mos. Thus it would seem that endogenous Emi1 is not sufficient to activate the CSF activity and that the activity of, or responsiveness to, endogenous Emi1 is regulated in a Mos-dependent fashion.

Given the involvement of Mos in the formation of the meiotic spindle^{11–14}, we propose that Mos-mediated modification of this spindle, perhaps at the kinetochores (where chromosomes attach), is required for Emi1 activity. If so, then blastomeres injected with Mos would arrest their division because Mos-induced changes in spindle structure increase its sensitivity to the effects of endogenous Emi1. By contrast, exogenous Emi1 might induce blastomere arrest because it compensates for the usual lack of sensitivity of the mitotic spindle to endogenous Emi1. This remains to be seen. But for now, it is satisfying to see that another piece in a 30-year-old puzzle is in place. ■

Nicholas S. Duesbery is in the Laboratory of Developmental Cell Biology, and George F. Vande Woude is in the Laboratory of Molecular Oncology, Van Andel Research Institute, 333 Bostwick NE, Grand Rapids, Michigan 49503, USA.
e-mail: george.vandewoude@vai.org

- Masui, Y. & Markert, C. L. *J. Exp. Zool.* **177**, 129–145 (1971).
- Reimann, J. D. R. & Jackson, P. K. *Nature* **416**, 850–854 (2002).
- Sagata, N., Watanabe, N., Vande Woude, G. F. & Ikawa, Y. *Nature* **342**, 512–518 (1989).
- Haccard, O. *et al. Science* **262**, 1262–1265 (1993).
- Gross, S. D., Schwab, M. S., Lewellyn, A. L. & Maller, J. L. *Science* **286**, 1365–1367 (1999).
- Bhatt, R. R. & Ferrell, J. E. Jr *Science* **286**, 1362–1365 (1999).
- Schwab, M. S. *et al. Curr. Biol.* **11**, 141–150 (2001).
- Reimann, J. D. *et al. Cell* **105**, 645–655 (2001).
- Colledge, W. H., Carlton, M. B., Udy, G. B. & Evans, M. J. *Nature* **370**, 65–68 (1994).
- Hashimoto, N. *et al. Nature* **370**, 68–71 (1994).
- Choi, T. *et al. Proc. Natl Acad. Sci. USA* **93**, 7032–7035 (1996).
- Verlhac, M.-H. *et al. Development* **122**, 815–822 (1996).
- Gross, S. D. *et al. Curr. Biol.* **10**, 430–438 (2000).
- Bodart, J. F. *et al. Dev. Biol.* (in the press).

Daedalus

Problems with niches

Why are there so many species? One answer is that each species has its own 'niche'. If two species occupy exactly the same niche, says the doctrine, one will drive the other extinct. Accordingly, the natural world is a mass of subtly different niches, each exploited by a specialized species. Hence, for example, a forest is a mass of different trees. Each has its own shape of leaf with its own chemical composition, discouraging the insects and other forest creatures that live in that niche from eating it. A 'standard leaf', optimizing the uptake of solar energy and common to many plants, has not evolved.

Sadly, a niche is left undefined. Yet the central claim is often made that just one species of a given type can inhabit one. To resolve the matter, Daedalus will try out 'artificial life'. He will allow a mix of programmed species to evolve in a computer. He recalls Robert Axelrod's famous work on how cooperation evolved. Many programs playing 'Iterated Prisoner's Dilemma' competed together, and the ensemble tested the way cooperation or competition survived or grew under 'darwinian' conditions. Similarly, if we could recognize a niche in practice, we could discover if niche theory made real sense.

So DREADCO information technologists are setting up a large number of closely related species in a big computer, and are setting them all to mutate and evolve simultaneously. There are two possible outcomes. The most successful species may drive all the others extinct, leaving it in sole command of the computer. Alternatively, the programs may diversify into a whole batch of different species, each fairly secure in some niche of program space. To make this work, a niche must be properly defined in computer terms.

Daedalus also recalls how the Soviet Union tried to identify, for example, the 'best' sewing machine. It could then be made in a huge factory for the whole Union. This idea came unstuck. By contrast, Western competition lets many small manufacturers compete, and each somehow survives in its own economic niche. Indeed, some analogy must relate an economic niche to a biological one. If Daedalus can spot it in practice, he will be able to find the ideal economic number of manufacturers, and compare it with the number found in free-market economics. With luck, the free market will turn out to be as closely ideal as freely evolved darwinian ecology is in the natural world.

David Jones

Obituary

César Milstein (1927–2002)

César Milstein began to study antibody diversity at a time when almost nothing was known about its molecular and genetic basis. As a logical step in that work, with Georges Köhler he devised the hybridoma technique for fusing cancer cells with antibody-producing cells to produce homogeneous (monoclonal) antibodies. This invention revolutionized wide areas of biomedicine, and Milstein and Köhler, together with Niels Jerne, were awarded the 1984 Nobel Prize in Medicine or Physiology.

Milstein was born into a Jewish family in Bahía Blanca, Argentina, and graduated from the University of Buenos Aires with a thesis on aldehyde dehydrogenase. In 1958, he won a scholarship to work at the University of Cambridge in Britain, where he did a second PhD, again in enzymology. There he met Fred Sanger, who had just won his first Nobel prize for determining the structure of insulin. Meeting Sanger was critical for Milstein's career. He returned to Argentina in 1961, but soon became dissatisfied with the political climate and, in 1963, took up Sanger's invitation to join the Medical Research Council's Laboratory of Molecular Biology (LMB), which had already attracted other top scientists from abroad such as Max Perutz, Aaron Klug and Sydney Brenner. It was on Sanger's advice that Milstein then began to study antibody structure and diversity: this was to be his life's work, with LMB in Cambridge his scientific home.

Milstein's hope was that the structural comparison of different antibodies would reveal the secret of antibody diversity. But things turned out to be more complicated. Soon after his arrival in Cambridge, it was discovered that antibody molecules contained variable (V) and constant (C) regions and the 'two genes, one polypeptide' hypothesis was proposed. According to this idea, separate genes encode V and C regions, and there are thousands of V genes in the germ line — the set of genes inherited by offspring. This was shockingly unorthodox, and other ideas were put forward. Among them, in 1966, was Brenner and Milstein's proposal of a mechanism that introduces 'somatic' (non-germline) mutations selectively into V segments of antibody genes. This proposal proved inadequate to explain primary antibody diversity. But it later became a leading model for the diversification of antibodies in immune responses, the main subject of Milstein's research for the latter part of his life.



Innovator in immunology

By the beginning of the 1970s, Milstein had become convinced that antibody structure could not explain the genetic basis of diversity. So he switched to work on the biosynthesis of antibodies and, with George Brownlee, on antibody messenger RNA. This required culturing antibody-producing myeloma cells and analysing them with molecular biological methods. It was this marriage of cellular and molecular techniques that led to the ensuing breakthroughs.

As a prelude, a biosynthetic precursor of parts of antibodies — the light chains — was discovered, providing early experimental evidence that a 'signal sequence' is required for protein transport across membranes. Most importantly, however, although it was initially only a sideline, the cell-fusion method was established in the laboratory to show that in hybrids of different myeloma cells antibody chains from both parental cells are produced in their original form. This was an early indication that, in antibody-forming cells, genes encoding the V and C regions were brought together, presumably by some translocation event. When Köhler arrived at LMB from the Basel Institute for Immunology, he brought with him an assay for identifying single antibody-producing cells that had been invented many years before by Jerne, director of the Basel Institute. With that, everything was in place for the great experiment.

"This time I can tell you something really interesting," César said with a smile, when I met him in 1975 at an obscure conference in San Remo, the kind of meeting he liked to attend. That something was the production of continuously growing, cloned cell lines secreting monoclonal antibodies of predetermined specificity, achieved by fusing myeloma

cells (contributing immortality and a high rate of antibody secretion) with antibody-producing cells from immunized mice (contributing antibody specificity). I felt the earth shaking.

This creation of hybridomas that produced antibodies of predetermined specificity was a dramatic advance. Antibody specificity and diversity could now be studied at the level of individual antibodies, and highly specific antibody reagents could be developed, allowing biological systems to be analysed with unprecedented resolution. Monoclonal antibodies have also become essential tools in medical diagnostics and are increasingly used in treating cancer and other diseases. A large part of the biotechnology industry is based on the monoclonal antibody technique.

Milstein contributed to these developments in manifold ways, seeing his invention as the beginning of a new era of 'antibody-based' molecules. But his real interest remained in antibody diversity, and he used the hybridoma technique to analyse how antibodies diversify during immune responses. This work showed the essential role of somatic mutation of antibody V genes in the affinity maturation of the antibody response. Mutations are introduced into these genes by a special hypermutation mechanism, as he and Brenner had speculated earlier. Now he tried to understand this mechanism through characterization of the pattern of mutations at the DNA level. He worked on this problem until the end of his life, and much of our knowledge of somatic hypermutation is based on the work of Milstein's Cambridge 'school'. Sadly, his wish to live long enough to see the solution of this problem was not fulfilled.

Milstein was awarded innumerable prizes and distinctions, but remained modest and easily approachable. He was not interested in academic power, always worked with a small group and liked to do experiments himself. He was insatiably curious, and his love of arguing about science — as well as about anything else in life — stemmed from that, and the belief and tradition that truth should be approached by dispute. He and his wife Celia liked to be with people, and to help them, and he had a special gift for friendship: his death on 24 March marks the loss not only of an outstanding scientist but of a man held in great affection by many.

Klaus Rajewsky

Klaus Rajewsky is at the Center for Blood Research, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA.
e-mail: rajewsky@cbr.med.harvard.edu

An unusual source of essential carotenoids

A yellow-faced vulture includes ungulate faeces in its diet for cosmetic purposes.

The rare Egyptian vulture (*Neophron percnopterus*) stands out among the Old World vultures (Family Accipitridae) because of its brightly ornamented head¹, which is coloured yellow by carotenoid pigments, and its practice of feeding on faeces. Here we show that Egyptian vultures obtain these pigments from the excrement of ungulates. To our knowledge, this is the first demonstration that faeces can be used as a source of carotenoids by a vertebrate.

Coprophagy is uncommon in birds, apart from the consumption of nestlings' faeces by parents in songbird species². The consumption of faeces is likely to be dangerous because they frequently contain high levels of parasites³. In addition, excrement is a poor source of the principal macronutrients (containing less than 5% protein and

under 0.5% fat, according to our analyses). However, the faeces of ungulates contain large amounts of intact carotenoids (see below), which have a pigmentary function⁴ and are also considered to be valuable micronutrients for vertebrates because of their antioxidant and immunostimulant properties⁵. Vertebrates are unable to synthesize essential carotenoids and must therefore obtain them by dietary means⁴.

Rotten flesh and bones, the typical diet of vultures, are poor sources of carotenoids. *N. percnopterus* may obtain these pigments from egg yolks or from insects, such as grasshoppers¹. The excrement of ungulates, however, is the most readily available and reliable source of carotenoids in areas covered by *N. percnopterus*. These vultures are often seen pecking at cow dung and eating the droppings of goats and sheep.

We used high-performance liquid chromatography⁶ to analyse the carotenoid content of ungulate faeces and to identify the carotenoids that are transported in the plasma and deposited in the yellow facial skin of *N. percnopterus* (Fig. 1). A single peak for lutein (which co-eluted with small quantities of zeaxanthin) was observed in all samples, accounting for over 95% of the total main carotenoid content. Lutein concentration in vulture skin (sampled from a freshly dead adult bird found in the wild) was $98 \mu\text{g g}^{-1}$, and a mean concentration of $6.5 \mu\text{g ml}^{-1}$ was found in plasma ($0.3\text{--}42.1 \mu\text{g ml}^{-1}$, $n = 196$). Mean lutein concentrations in faeces were $35.7 \mu\text{g g}^{-1}$ in cow dung, $185.8 \mu\text{g g}^{-1}$ in sheep droppings, and $36.6 \mu\text{g g}^{-1}$ in goat droppings (two samples were analysed from each species of ungulate).

Additional evidence, albeit indirect, to support our idea that Egyptian vultures obtain carotenoids from faeces is provided by the fact that higher plasma carotenoid concentrations were evident in individual birds from areas with greener pastures, where free-grazing cattle are more abundant (Fig. 2a).

We confirmed that *N. percnopterus* obtains carotenoids from cow dung by conducting a food trial on four adult birds at Jerez Zoo, Spain. These birds had been fed a cow-meat diet for several months and only had traces of lutein (less than $1 \mu\text{g ml}^{-1}$) in their plasma when the trial began. They were fed exclusively on an unlimited diet of fresh cow dung for 10 days, of which they consumed about 1.3 kg in total. All four birds showed significantly higher plasma lutein levels at the end of this trial than at the beginning (Fig. 2b).

The cow dung used in the food trial



Figure 1 The Egyptian vulture has earned the nickname in Spain of 'churretero' or 'moñiguero', meaning 'dung-eater'.

contained 4% semi-decomposed protein (determined using the Kjeldahl method⁷) and 0.38% fat (determined using the Soxhlet method⁷), and is therefore a poor source of major nutrients. However, the high concentrations of carotenoids in cow dung mean that the adoption of a coprophagous habit by *N. percnopterus* ensures the intake of an excess of dietary carotenoids, well above the levels needed for healthy physiological function.

This excessive consumption of pigment suggests a possible evolutionary mechanism for the development of carotenoid-dependent ornaments in *N. percnopterus*. Surplus carotenoids would diffuse passively to the skin⁸ and the resulting yellow coloration could have become a useful signal in mating displays⁹, for advertising dominant status¹⁰, or both.

For the signal to be reliable, however, it must be associated with costs that can only be met by high-quality individuals¹¹. One potential cost of eating excrement is that it may expose the consumer to gastrointestinal parasites³; another is the time and effort involved in searching for food with negligible nutritional content.

J. J. Negro*, J. M. Grande*, J. L. Tella*, J. Garrido†, D. Hornero†, J. A. Donazar*, J. A. Sanchez-Zapata‡, J. R. Benítez§, M. Barcell§

*Department of Applied Biology, Estación Biológica de Doñana, CSIC, Pabellón del Perú, Avda María Luisa, s/n 41013 Seville, Spain
e-mail: negro@ebd.csic.es

†Food Biotechnology Department, Instituto de la Grasa, CSIC, Avda Padre García Tejero 4, 41012 Seville, Spain

‡Area de Ecología, Departamento de Biología Aplicada, Universidad Miguel Hernández, 03312 Alicante, Spain

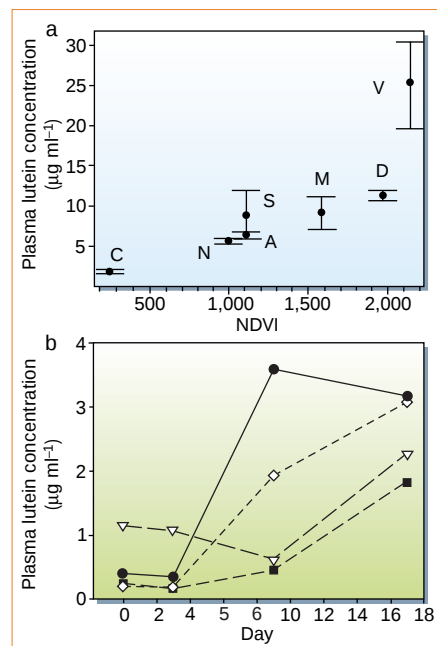


Figure 2 Plasma lutein concentrations in Egyptian vultures (*Neophron percnopterus*). **a**, Mean concentrations of lutein $\pm$ s.e.m. in the plasma of wild Egyptian vultures. Sampling areas (all in Spain; sample sizes in parentheses): C, Canary Islands (22); N, Navarra (72); A, Aragón (77); S, Segovia (4); M, Menorca (4); D, Andalucía (11); V, Vizcaya (6). Lutein concentrations were measured by spectrophotometry¹². The mean lutein concentration in each population correlates with the local values of a vegetation index (NDVI), which is an indicator of photosynthetic production¹³ ($r = 0.93$, $P < 0.01$, $n = 7$). **b**, Plasma lutein concentration ($\mu\text{g ml}^{-1}$) of four captive adult Egyptian vultures measured by high-performance liquid chromatography⁶. Before day 0, the birds were fed on cow meat. On days 0–9, they had access to cow dung only. From day 9 onwards, the birds were returned to the meat diet and no dung was available. Plasma lutein concentrations increased significantly over this time (repeated-measures ANOVA, $F_{3,9} = 4.799$, $P = 0.0298$).

§Zoo de Jerez, Taxdirt s/n 11404 Jerez de la Frontera, Spain

- Mundy, P., Butchart, D., Ledger, J. & Piper, S. *The Vultures of Africa* (Academic, London, 1992).
- Hurd, P. L., Weatherhead, P. J. & McRae, S. B. *Behav. Ecol.* **2**, 69–76 (1991).
- Hutchings, M. R., Gordon, I. J., Kyriazakis, I. & Jackson, F. *Anim. Behav.* **62**, 955–964 (2001).
- Brush, A. H. *FASEB J.* **4**, 2969–2977 (1990).
- Britton, G. *FASEB J.* **9**, 1551–1558 (1995).
- Mínguez-Mosquera, M. I. & Hornero-Méndez, D.

- J. Agric. Food Chem.* **41**, 1616–1620 (1993).
- Vogel, A. I. *Elementary Practical Organic Chemistry* (Logmans, Green & Co., London, 1957).
- Negro, J. J., Tella, J. L., Blanco, G., Forero, M. G. & Garrido-Fernández, J. *Physiol. Biochem. Zool.* **73**, 97–101 (2000).
- Hill, G. E. *Nature* **350**, 337–339 (1991).
- Rohwer, S. A. *Evolution* **29**, 593–610 (1975).
- Zahavi, A. & Zahavi, A. *The Handicap Principle* (Oxford Univ. Press, Oxford, 1997).
- Tella, J. L. et al. *Physiol. Zool.* **71**, 708–711 (1998).
- Díaz-Delgado, R. & Pons, X. *Forest Ecol. Manag.* **147**, 67–74 (2001).

Competing financial interests: declared none.

Biodiversity

Invasions by marine life on plastic debris

Colonization by alien species poses one of the greatest threats to global biodiversity¹. Here I investigate the colonization by marine organisms of drift debris deposited on the shores of 30 remote islands from the Arctic to the Antarctic (across all oceans) and find that human litter more than doubles the rafting opportunities for biota, particularly at high latitudes. Although the poles may be protected from invasion by freezing sea surface temperatures, these may be under threat as the fastest-warming areas anywhere² are at these latitudes.

Like humans, marine organisms are now experiencing unparalleled availability, distribution and duration of transport. Floating debris is the most common sea-going transport system^{3–5} and is responsible for the widespread distribution of many marine animals that use it to hitch a ride. Natural debris such as volcanic rock, or pumice, and wood have always carried

organism propagules. But there has recently been an explosive increase in anthropogenic debris as a result of massive amounts of plastic entering the oceans — for example, the amount of debris doubled from 1994 to 1998 around the coastline of the United Kingdom⁶, and in parts of the Southern Ocean it increased 100-fold during the early 1990s⁷.

I examined about 200 items of debris washed ashore on each of 30 islands (Fig. 1), which were scattered over a geographical range extending from Spitsbergen in the Arctic to Signy Island in the Antarctic. I found that 20–80% of this debris was anthropogenic (man-made) in origin (Fig. 2a), of which the highest proportion was in the Southern Ocean (as land there has no forests, there is scarcely any local natural debris). There was generally less anthropogenic debris at low latitudes in the Southern Hemisphere than at equivalent northern latitudes, presumably because there are fewer people there and more ocean.

Many types of animal use marine debris as a mobile home, particularly bryozoans, barnacles, polychaete worms, hydroids and

molluscs (in order of abundance), and I found that among the most numerous were animals with cosmopolitan distributions. The successful natural dispersal of species with longer-lived planktonic larvae, such as the bryozoan *Membranipora*, is unsurprising, but most of the other colonizing animals, such as the hydroid *Halecium*, have brooded or brief-duration larvae and no obvious means of dispersal. Oceanic debris therefore offers a major opportunity for dispersal of such species.

Another opportunity for invasion of new habitats by alien organisms is presented by their adherence to ships' hulls. Compared with boats, however, man-made debris is longer lasting, more pervasive and travels more slowly, factors that could favour the survival of colonists.

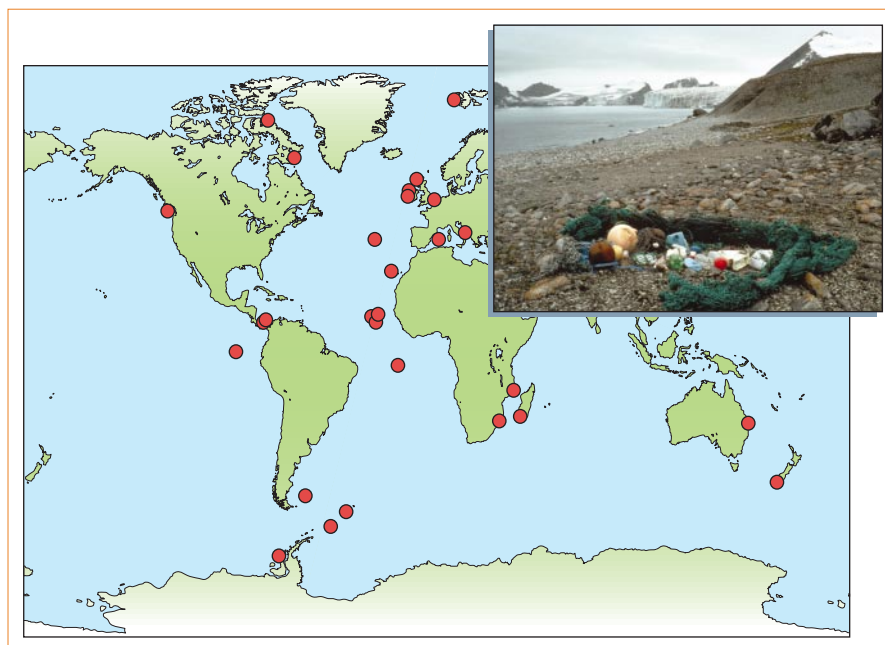


Figure 1 A global picture of shore debris. Dots show the locations of the sampling sites. Inset, typical examples of raft debris washed ashore. The movement of such debris has increased the propagation of colonizing fauna, threatening biodiversity in many regions.

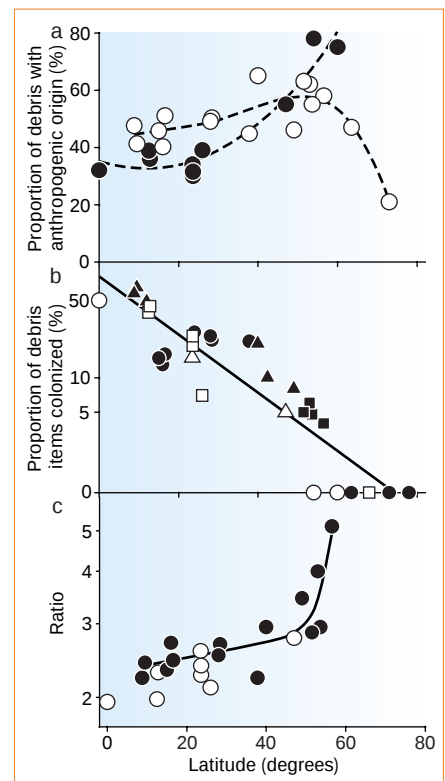


Figure 2 Colonization of man-made and natural debris by marine organisms at different latitudes. **a**, Proportion of man-made debris found offshore at 30 remote islands (Fig. 1); debris is classed as either anthropogenic (mainly plastic) or natural (mostly wood, but not lumber); $n \approx 200$ for each point. Open symbols, islands in the Northern Hemisphere; filled symbols, islands in the Southern Hemisphere. **b**, Variation with latitude, hemisphere and remoteness of island shorelines in the proportion of marine debris of each type that was colonized by fauna. Symbols represent the distance of each island from the continental mainland: circles, hundreds of kilometres; triangles, tens of kilometres; squares, less than 10 km. Fitted regression line has associated $r^2 = 72.7$, and significance by ANOVA: $F = 85$, $P \leq 0.001$. **c**, Variation with latitude and hemisphere in the ratio of propagules on non-anthropogenic to those on anthropogenic debris on island shores. Data are calculated from the proportional colonization by different types of debris shown in **a** at the latitudes given in **b**. The best-fit curve is exponential and has associated $r^2 = 0.74$ and significance by ANOVA: d.f. = 4, $F = 10.25$, $P = 0.002$.

It has been proposed⁸ that floating debris might be colonized by marine organisms in a cline from the poles to the tropics, and that the density and type of recruit might also vary with latitude. The global data set that I collected from surveying marine debris on 30 island shores not only shows a striking increase with latitude in the proportion of debris that is man-made (Fig. 2a), but also in the potential for transporting organisms (Fig. 2b). Although this survey leaves many unknowns, including time spent at sea, drift path, route, proximity of larval supply sources and even point of origin, a strong latitudinal trend still emerges (Fig. 2b). The strength of this pattern, the hemisphere symmetry and the agreement with other results⁹ suggests that the relationship is robust.

Surprisingly, the distance from the mainland to each island does not seem to influence the proportion of debris colonized. Latitude is a good predictor of debris colonization up to about 60°, beyond which no samples were found to carry fauna. Possible explanations include increased agitation of the raft by wave action or ice scour, the lack of pole-bound currents (in Antarctica) or increased ultraviolet radiation mediated by decreased stratospheric ozone, but persistently low temperatures are the most likely.

If freezing temperatures are the main barrier to invasion by marine-borne organisms, then polar warming (a recent model predicts a temperature rise of more than 2 °C over the next 100 years in the Southern Ocean¹⁰) could alleviate this constraint. Whereas the geologically young Arctic Ocean is still being invaded by new species¹¹, the Antarctic Southern Ocean, which has been isolated by the circumpolar current for at least 25 million years, is largely populated by endemic species. Invasion from rafting biota could therefore be ecologically more significant in the Southern Ocean, where the highest proportion of man-made debris is found, making Antarctic waters relatively (not absolutely) more susceptible.

I estimate that rubbish of human origin in the sea has roughly doubled the propagation of fauna in the subtropics and more than tripled it at high (>50°) latitudes (Fig. 2c), increasing the potential for alien invasion and adding to the problems already created by sea-borne plastic materials in the form of injuries and mortality among marine mammals and birds^{7,12,13}.

David K. A. Barnes

British Antarctic Survey,
Natural Environment Research Council,
High Cross, Madingley Road, Cambridge,
CB3 0ET, UK, and Department of Zoology and
Animal Ecology, University College Cork, Cork,
Ireland

e-mail: dkab@bas.ac.uk

1. Goldberg, E. D. *Chem. Ecol.* **10**, 3–8 (1995).
2. Quale, W., Peck, L. S., Peat, H., Ellis-Evans, J. C. & Harrigan, P. R. *Science* **295**, 645 (2002).
3. Thresher, R. E. & Brothers, E. B. *Evolution* **39**, 878–887 (1985).
4. Highsmith, R. C. *Mar. Ecol. Prog. Ser.* **25**, 169–179 (1985).
5. Jokiel, P. L. *Endeavour* **14**, 66–73 (1990).
6. *British Ecological Society Bulletin* Annual Report (1999).
7. Walker, T. R., Reid, K., Arnould, J. P. Y. & Croxall, J. P. *Mar. Poll. Bull.* **34**, 61–65 (1997).
8. Barnes, D. K. A. & Sanderson, W. A. in *Proc. 11th Int. Bryozool. Assoc. Conf.* (eds Herrera Cubilla, A. & Jackson, J. B. C.) 154–160 (Smithsonian Tropical Res. Inst., Panama, 2000).
9. Winston, J. E., Gregory, M. R. & Stevens, L. M. in *Marine Debris: Sources, Impact and Solutions* (eds Coe, J. M. & Rogers, D. B.) 81–97 (Springer, New York, 1997).
10. Murphy, J. M. & Mitchell, J. F. B. *J. Climate* **8**, 469–514 (1995).
11. Dunton, K. *Trends Ecol. Evol.* **7**, 185–189 (1992).
12. Gregory, M. R. *Ocean Coastal Manag.* **42**, 603–615 (1999).
13. Bonner, W. N. & McCann, T. S. *Br. Antarct. Surv. Bull.* **57**, 73–77 (1982).

COMMUNICATIONS ARISING

Spintronics

Spin accumulation in mesoscopic systems

In spintronics, in which use is made of the spin degree of freedom of the electron, issues concerning electrical spin injection and detection of electron spin diffusion are fundamentally important. Jedema *et al.* describe a magneto-resistance study in which they claim to have observed spin accumulation in a mesoscopic copper wire¹, but their one-dimensional model ignores two-dimensional spin-diffusion effects, which casts doubt on their analysis. A two-dimensional vector formalism of spin transport^{2,3} is called for to model spin-injection experiments, and the identification of spurious background resistance effects is crucial.

As predicted⁴, spin-polarized electrons injected from a ferromagnetic film, F1, into

a non-magnetic metal, N, (Fig. 1a) accumulate in N when the injection rate, I_M , is greater than the spin-relaxation rate, $1/T_2$. Spin accumulation (equivalent to the non-equilibrium magnetization, $\vec{M}$) in the volume of N near F1 is proportional to the difference of up- and down-spin sub-band electrochemical potentials, $\vec{M} = I_M T_2 / \text{vol}$, which is proportional to $\mu_{\text{up}} - \mu_{\text{down}}$, and decays exponentially in N on the length scale of the classical spin depth, δ_N (refs 2,3). The spin-coupled voltage, V_s , detected at a film F2 (for sample dimensions $< \delta_N$) is

$$V_s/I = R_s = (\eta^2 \rho_N \delta_N^2) / \text{vol} \quad (1)$$

where ρ_N is the resistivity of N, and η is the polarization injection efficiency⁵. Spin accumulation can be detected as a resistance change $\Delta R = 2R_s$ when magnetic fields in the x–y plane change the magnetization orientations $\vec{M}_1$ and $\vec{M}_2$ from parallel to antiparallel^{5,6}.

In the geometry and model described by Jedema *et al.*, permalloy films Py1 and Py2 are located on the vertical arms of a copper (van der Pauw) cross and have edges at $y = \pm L/2$. Spins are injected at the Py1–Cu edge, diffuse into the copper, and may be detected at Py2. For the device with $L = 250$ nm at $T = 4$ K, the expected magnitude of ΔR is found from equation (1) by using¹: $\text{vol} = d_N \times W_N \times L$, $d_N = 50$ nm, $W_N = 100$ nm, $\rho = 1.5 \times 10^{-6}$ Ω-cm, $\eta = 40\%$ (ref. 7) and (from $\tau/T_2 = 0.001$; refs 5, 6) $\delta_N = \lambda_N = 1.0$ μm. The estimate $\Delta R = 1.9$ Ω is 10^3 times greater than the measured value¹, $\Delta R = 1.5 \times 10^{-3}$ Ω.

To account for this discrepancy, Jedema *et al.* claim that a ‘resistance mismatch’ at the F–N interface diminishes the polarization in proportion to

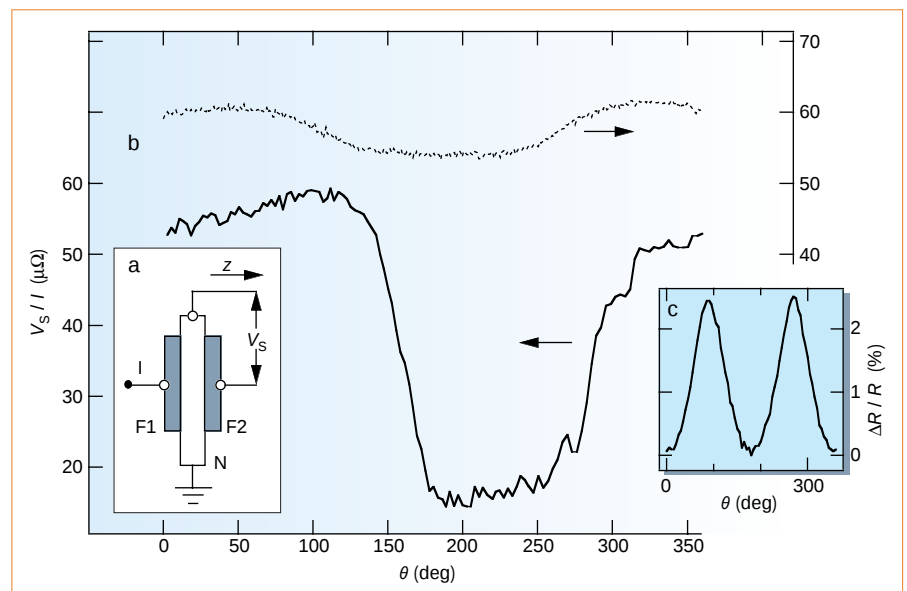


Figure 1 Background resistance and electron spin-diffusion effects in an all-metal mesoscopic spin valve. **a**, Cross-section of pedagogical spin-injection device. **b**, $\cos \theta$ angular symmetry of spin accumulation. Field **H** is rotated in the x–y plane (see **a**). **c**, $\cos 2\theta$ angular symmetry of anisotropic magneto resistance, measured in a 100-nm-thick permalloy film at 76 K.

$M = (\delta_N \rho_N / \delta_F \rho_F)(1 - \alpha_F^2)$ (ref. 8), where α_F is the polarization of bulk permalloy. Although no current flows in F2 in an open-circuit voltmeter measurement, a 'mismatch' is also invoked at the Cu-Py2 interface, and the model proposed by Jedema *et al.* for ΔR (for dimensions $< \delta_N$) differs from that given in equation (1) by the factors $M^{-1}(M+1)^{-1} \approx 10^{-3}$.

However, this assertion relies on the erroneous assumption that the intrinsic interface resistance, R_i , is perfect, $R_i = 0$. In theory^{3,8}, an F-N interface is characterized by $r_N = (\delta_N \rho_N)$ and $r_F = (\delta_F \rho_F)$, relative to R_i . A 'resistance mismatch' is important only if $R_i \ll r_N, r_F$. This issue is analogous with spin injection at an F-semiconductor interface¹⁰. Efficient spin injection is achieved when a barrier, with resistance as small as r_s or r_F , mediates transport. For F-N interfaces, a contact resistance, R_C , exists at metal-metal interfaces that are formed by lift-off processes, regardless of ion-mill cleansing. R_C dominates interfacial transport, $R_C = R_i \geq r_i \approx 10^{-11} \Omega\text{-cm}^2$, and the interface parameter η , rather than α , characterizes spin injection.

I therefore question the validity of the model proposed by Jedema *et al.*. If the data are related to spin accumulation, the discrepancy between theory and experiment is derived from a misapplication of one-dimensional Valet-Fert theory to a two-dimensional system, errors that are independent of F-N interface contact resistance. First, the resistance per square of copper is much less than that of the permalloy, $R_{(\text{sq,Cu})} = 0.3 \Omega < R_{(\text{sq,Py})} = 2.0 \Omega$, and the copper wire that spans Py1 shunts the imposed bias current. The current density injected across the F1-N interface is one-seventh of the modelled value.

Second, the film Py2 supposedly measures a signal that is related to "...densities of the spin-up and spin-down electrons in the centre of the cross...", which I question. Spin detection is an interfacial effect^{3,8} and depends on the spin densities at the N-F2 interface. Detection is averaged over the width of Py2 and ΔR is diminished by roughly half.

Third, spin diffusion in the copper is isotropic. Only half of $\vec{M}$ diffuses towards the centre of the cross, and one-third of this population diffuses down each of the remaining arms. For $L < \delta_N$, the spin density near Py2 is diminished by a factor of one-sixth. Together, these errors give factors of 10–100 that vary with L , a fact that undermines the analysis of Jedema *et al.*

The observed variations in $\Delta R/R$ of 0.1–5.0% are extremely small. As the model of Jedema *et al.* is sensitive to ρ_F , and ΔR is proportional to ρ_F^2 , it cannot distinguish between anisotropic magneto resistance, $\delta \rho_F / \rho_F = 2.5\%$, and spin injection. Anisotropic magneto resistance can be

measured by determining the angular symmetry of R as an external field is rotated in the sample plane. For the example shown in Fig. 1b, the coercivities of F1 and F2 of the spin-transistor sample^{5,6} are $\mathbf{H}_{C,1} = 10$ Oe and $\mathbf{H}_{C,2} = 21$ Oe. After applying a field $\mathbf{H} = 100$ Oe at $\theta = 0^\circ$, $\mathbf{H}$ is reduced to $\mathbf{H}_{C,1} < \mathbf{H} = 15$ Oe $< \mathbf{H}_{C,2}$, such that $\mathbf{M}_2$ remains oriented along $\theta = 0^\circ$, but $\mathbf{M}_1$ can follow $\mathbf{H}$ through a 360° rotation. Referring to the lower trace (left axis), $R = V_s/I$ is minimum after $\theta \approx 180^\circ$ when $\mathbf{M}_1$ and $\mathbf{M}_2$ are antiparallel, and maximum near $\theta \approx 0^\circ$ when $\mathbf{M}_1$ and $\mathbf{M}_2$ are parallel. This $\cos \theta$ symmetry is characteristic of spin accumulation. Next, $\mathbf{H} = 100$ Oe is applied and then reduced to $\mathbf{H}_{C,1} < \mathbf{H}_{C,2} < \mathbf{H} = 40$ Oe, so $\mathbf{M}_1$ and $\mathbf{M}_2$ are always roughly parallel and V_s (upper trace, right axis) is roughly constant. By contrast, anisotropic magneto resistance has two maxima in a 360° rotation (Fig. 1c).

I question the general validity of 'resistance mismatch' at F-N interfaces; the one-dimensional model is probably incorrect, and the results of Jedema *et al.*¹ may require a new interpretation.

Mark Johnson

Naval Research Laboratory,
Washington DC 20375, USA
e-mail: mjbooj@anvil.nrl.navy.mil

1. Jedema, F. J., Filip, A. T. & van Wees, B. J. *Nature* **410**, 345–348 (2001).
2. Johnson, M. & Silsbee, R. H. *Phys. Rev. Lett.* **55**, 1790–1793 (1985).
3. Johnson, M. & Silsbee, R. H. *Phys. Rev. B* **37**, 5312–5325 (1988).
4. Silsbee, R. H. *Bull. Mag. Res.* **2**, 284–285 (1980).
5. Johnson, M. *J. Appl. Phys.* **75**, 6714–6719 (1994).
6. Johnson, M. *Phys. Rev. Lett.* **70**, 2142–2145 (1993).
7. Soulen, R. J. *et al. Science* **282**, 85–88 (1998).
8. Johnson, M. & Silsbee, R. H. *Phys. Rev. B* **35**, 4959–4972 (1987).
9. Johnson, M. & Silsbee, R. H. *Phys. Rev. Lett.* **60**, 377 (1988).
10. Rashba, E. I. *Phys. Rev. B* **62**, 16267–16270 (2000).

Jedema et al. reply — Johnson suggests that our results¹ are not due to spin transport but rather are caused by spurious effects, in particular the anisotropic magneto resistance of the ferromagnetic contacts. However, our experiment was designed explicitly to eliminate any magneto-resistance effects that might arise from the ferromagnetic contacts.

We used the well known principle of four-terminal measurement. No current flows through the ferromagnetic detector contact, so its (magneto) resistance does not affect the measurement. Similarly, the (magneto) resistance of the ferromagnetic injector contact is not relevant, because any voltage that develops across it will not affect

the current that is sent through it. The only region that could then possibly give rise to a (magneto) resistance is the central part of the copper cross, but explicit measurements confirm that this region does not produce any magneto resistance. So, having eliminated all other possible contributions, the only mechanism left that can give rise to a signal is spin accumulation. In view of this, and of the consistency of our measurements and analysis^{1,2}, the suggestion that we cannot distinguish between spin accumulation and anisotropic magneto resistance is unjustified.

Johnson also points out that if we have observed spin accumulation, then our conclusion that the spin polarization of the current is low (1–2%) must be wrong — claiming that it should be considerably higher (around 40%). Our detailed theoretical model² reveals that our conclusion does not crucially depend on specific details of the injection mechanism. Moreover, we did not invoke conductivity mismatch³. It is an integral part of the description, in which the boundary conditions imposed by the contacts are included in a straightforward way.

Consistent treatment of the injector and detector contacts^{4,5} is missing from Johnson's description, which therefore cannot be applied directly to the case of transparent contacts. Moreover, we did not observe any interface resistance between ferromagnetic and non-magnetic regions, as Johnson suggests, which could be distinguished from the bulk diffusive resistance. There is also no physical reason why there should be such a large resistance at a (disordered) interface between metals.

Our experiments enabled us to observe spin accumulation in its purest form, without any contribution from spurious effects. The low spin polarization of the injected current can be explained by proper modelling of the entire system, including the contacts.

F. J. Jedema, A. T. Filip, B. J. van Wees

Department of Applied Physics and Materials
Science Centre, University of Groningen,
9747 AG Groningen, The Netherlands
e-mail: b.j.van.wees@phys.rug.nl

1. Jedema, F. J., Filip, A. T. & van Wees, B. J. *Nature* **410**, 345–348 (2001).
2. Jedema, F. J., Nijboer, M. S., Filip, A. T. & van Wees, B. J. *J. Supercond.* **15**, 27–35 (2002).
3. Schmidt, G., Ferrand, D., Molenkamp, L., Filip, A. T. & van Wees, B. J. *Phys. Rev. B* **62**, R4790 (2000).
4. Fert, A. & Lee, S. F. *Phys. Rev. B* **53**, 6554–6565 (1996).
5. Herschenfeld, S., & Zhao, L. Z. *Phys. Rev. B* **56**, 3296–3305 (1997).

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com) or on request from nature@nature.com

Insights into phase transition kinetics from colloid science

Valerie J. Anderson*† & Henk N. W. Lekkerkerker*

* Van't Hoff Laboratory for Physical and Colloid Chemistry, Debye Institute, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Colloids display intriguing transitions between gas, liquid, solid and liquid crystalline phases. Such phase transitions are ubiquitous in nature and have been studied for decades. However, the predictions of phase diagrams are not always realized; systems often become undercooled, supersaturated, or trapped in gel-like states. In many cases the end products strongly depend on the starting position in the phase diagram and discrepancies between predictions and actual observations are due to the intricacies of the dynamics of phase transitions. Colloid science aims to understand the underlying mechanisms of these transitions. Important advances have been made, for example, with new imaging techniques that allow direct observation of individual colloidal particles undergoing phase transitions, revealing some of the secrets of the complex pathways involved.

Figure 1 illustrates three types of phase diagram. The first (Fig. 1a) is a simple system of hard spheres. Introducing attractions results in three-phase equilibria, as in atomic systems such as argon (Fig. 1b). With shorter-range attractions the gas–liquid (or fluid–fluid) equilibrium becomes metastable (Fig. 1c). This is often observed in protein systems. One might assume equilibrium diagrams show the complete picture, with perhaps the initial distance from the phase boundary (indicative of the distance to equilibrium) controlling the speed with which the equilibrium phases are attained; however, this is the exception rather than the rule.

Hard-sphere colloids suspended in a solvent provide an excellent illustration of the difficulties involved in understanding the equilibrium states and the mechanisms by which systems evolve. Entropy considerations predict that these systems will form crystals if the volume fraction is increased. Above the ‘freezing’ volume fraction, $\phi_f = 0.494$, it is entropically favourable if some spheres are in a crystal, but above the ‘melting’ volume fraction, $\phi_m = 0.545$, all spheres should be in a crystal (Fig. 1a). However this is not always the situation found experimentally, either because the conditions the theory assumes (low polydispersity, for example) are not satisfied, or the dynamics of the system have dictated a different structure (note that we cannot say that this structure is the equilibrium structure—if entropy favours crystal formation then a crystal will form eventually, however the system may remain in the less-favoured state for a significant amount of time). This crystallization process is often interpreted within the familiar framework of nucleation and growth. There has been renewed interest in this mechanism recently, both with simulations¹ and experimentally^{2,3}: by monitoring the individual particles undergoing the transition it is possible to evaluate and improve the model.

The classical theory predicts that the free energy cost, ΔG , of forming a nucleus of radius r is:

$$\Delta G = \frac{4\pi r^3 \rho \Delta\mu}{3} + 4\pi r^2 \gamma$$

where γ is the surface free energy, ρ is the density of the bulk liquid, and $\Delta\mu$ is the chemical potential difference between the bulk solid and bulk liquid. For a large enough radius the (favourable) volume term begins to dominate the (unfavourable) surface term, the nucleus becomes stable, and may grow.

Directly observing crystal nucleation is difficult, often because once crystal nuclei are big enough to be seen they are well beyond the critical stage. Extensive simulation work has been done but there are still discrepancies, for example, in the prediction of the nucleation rate. Today, advances in microscopy and computational technology mean that particles in crystallizing samples can be tracked, effectively letting us run time ‘backwards’ from the large, growing crystals to the earlier, critical times to examine the critical nuclei^{2,3}. This allows assumptions and predictions from simulations and experiments to be compared (Fig. 2). One premise that may be disputed is that of a spherical nucleus. The critical nucleus may be anisotropic with different free energy costs associated with different faces (this would have repercussions on the growth rate as different faces are likely to grow at different rates). Using confocal microscopy, Gasser *et al.*³ have observed elliptical nuclei (Fig. 2b). A second discrepancy between simulations and experiment occurs in the value of the surface tension: it is almost seven times lower³ than found by simulations¹. However, a small amount of charge in the system of ref. 3 means that the freezing and melting transitions are observed at $\phi_f = 0.38$ and $\phi_m = 0.42$, so it would be interesting to determine whether this small charge can also affect the surface tension.

The packing of the spheres within the crystal also requires attention. Although the most stable is face-centred cubic packing⁴ (f.c.c.), the Ostwald rule of stages states that the transition from one

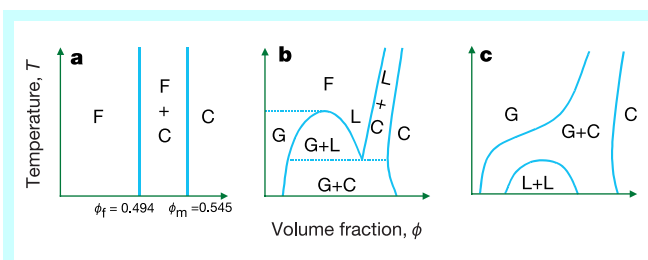


Figure 1 A wide range of phase diagrams occurs naturally. **a**, For a purely hard-sphere system the phase diagram shows only fluid (F) and crystal (C) phases. **b**, Atomic systems are often modelled by hard spheres with long-range attractions. This leads to equilibrium between gas (G), liquid (L) and crystal phases. **c**, In cases where the attraction is short-range, as in protein systems (important in physiology), equilibrium between gas and crystal is found, but the liquid–liquid transition becomes metastable.

† Present address: School of Chemistry, University of Bristol, Cantocks Close, Bristol, BS8 1TS, UK.

phase to another proceeds via the intermediate metastable states. Alexander and McTague⁵ argued that the first step would be body-centred cubic packing (b.c.c.), but this is rarely seen, even as an intermediate. Instead, the crystal usually shows random hexagonal close-packing (h.c.p.), even in the most ideal conditions⁶, although the degree of randomness appears to depend on growth rate⁷. This anneals to the more stable f.c.c., but because the energy difference between the h.c.p and f.c.c. structures is small, the driving force is also small and the step may be slow, depending also on factors such as crystal size⁸.

Pusey and van Megen⁹ clearly observed that above a certain volume fraction ($\phi_g = 0.58$) an amorphous glassy phase appeared that did not crystallize over several months. The kinetic arrest found in glass formation occurs in many systems^{10,11} and its understanding is an important scientific goal. Important in understanding the glass transition is the notion of spatial heterogeneities in the sample, that is, distinct relaxing domains or regions where particles move cooperatively. Microscopy has proven invaluable in looking directly at particle movement in colloidal samples, allowing characterization of the dynamics and structure in glassy systems^{12–15}. Until 1997 it was generally thought that $\phi_g = 0.58$ was an intrinsic glass transition volume fraction for hard-sphere colloidal suspensions, but then samples with high volume fractions, which fail to crystallize even after a year on Earth, were seen⁶ to crystallize fully in less than two weeks in microgravity (order of magnitude, $10^{-6}g$). This has also

been corroborated in experiments on centrifuged samples¹⁶ and indicates that even a small amount of a uniaxial stress can jam a system such that crystallization is not observed, as illustrated in Fig. 3.

These findings illustrate the complications found in the simple case of entropy-driven crystallization of hard spheres, where the phase diagram is a straightforward crystal–fluid equilibrium, but the complementary techniques of microscopy and computer simulation have led to advances in the understanding of this phase transition. This also aids the interpretation of the same mechanism in less straightforward systems.

Colloid–polymer systems

The fluid–solid transition can be complicated, but transitions involving three-phase equilibria or metastable states, as found in more familiar atomic and molecular systems, can be even more tortuous, as can be inferred from the increased complexity of the phase diagrams. An analogous situation can be produced in colloidal systems where the addition of polymer is used to produce an attraction between the particles (Fig. 4a). Depending on the relative sizes of the polymer and the colloid, such systems not only show the gas–crystal coexistence described above, but also gas, liquid and crystal coexistence, leading to remarkable phase behaviour.

The depletion interaction that is induced by adding polymer to a colloidal system effectively widens the region over which phase separation is expected (illustrated in Fig. 4b, c). An advantage of these systems is that the range and depth of the potential are controlled by the size and volume fraction of the polymer respectively. For a given colloid size, two distinct situations can be resolved, depending on the polymer size (Fig. 4b, c). The polymer concentration C_p is akin to inverse temperature: compare Fig. 4b, c with Fig. 1b, c (which illustrate predictions for atomic and protein systems). Because the range of the interaction can be tuned, a variety of phase diagrams can be realized. This allows the study of transitions both in systems with long-range attractions and those with very short-range attractions where the phase diagrams predict intriguing metastable states.

Although it is possible to predict the equilibrium phases of such systems, they also show frustrating behaviour; for many systems a colloidal crystal is expected but not achieved. In some instances not even the colloid-rich phase is attained, the system being trapped in a

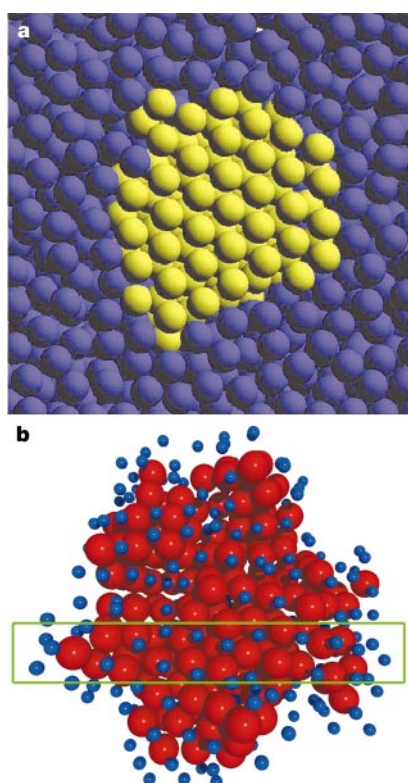


Figure 2 Nucleation events in simulations and in experiments. **a**, Simulations provide insights into the structure, shape and packing of critical nuclei, as this image from the work of Auer and Frenkel¹ illustrates. Until recently, this was the main technique with which such events could be observed; however, advances in imaging technology now allow critical nuclei to be imaged in three dimensions. **b**, The crystallization of poly-methyl methacrylate colloids was followed in real time by Gasser *et al.*³, and crystal nuclei were identified, allowing comparisons to be made with the results of simulations and theory. Reproduced with permission from ref. 3. Copyright (2002) American Association for the Advancement of Science.

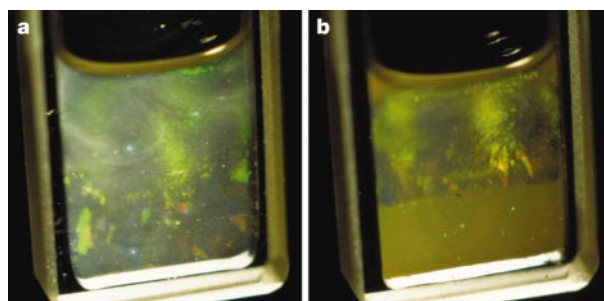


Figure 3 Suppression of crystallization by a small amount of uniaxial stress¹⁶. The samples both show poly-methyl methacrylate spheres in a solvent where there is only a very small density difference between colloid and solvent. The sample in **b**, however, shows for fewer crystal regions than the sample in **a**. This difference in degree of crystallization was caused by applying a small amount of stress (by centrifuging the sample) to the sample in **b**. Although the stress was applied only for a period of hours and the samples were then left to crystallize for over a year, the experiment demonstrates how dramatically stress can jam a system. Reproduced with permission from ref. 16. Copyright (2000) American Chemical Society.

gel-like phase instead. Much depends on the initial conditions, in this case, polymer concentration and colloid volume fraction. Although slightly changing one condition may not alter which equilibrium phases are expected, it may significantly affect whether or not they are reached (Fig. 4d). Thus, to predict the transition mechanisms of both colloid and colloid–polymer systems reliably is difficult.

Predictions from phase diagrams and free energy

One illustration of the importance of the kinetics and mechanisms to an actual experiment is recent work on the diffusive growth of crystals from polydisperse spheres¹⁷. The essence is that because smaller spheres move fastest, they arrive at the crystal surface first and the resulting crystal has more small particles than one might otherwise expect. Although it is predicted that polydisperse systems will fractionate in order to form crystals^{18–20}, with the simplest case comprising a crystal and coexisting fluid with different size distributions and thus different polydispersities, this is a thermodynamic route to fractionation; the fractionation via diffusive growth, described above, is a dynamic route. (We note that the thermodynamic fractionation requires the crystals to be made up of the larger particles.) Recent work indicates that polydispersity can greatly affect the crystal nucleation rate because of the increase in the crystal–fluid surface free energy²¹, leading to intriguing predictions concerning, for example, the variation in the density of the critical nucleus²². Polydispersity may also play a part in the slowing of the kinetics on the approach of the glass transition²³.

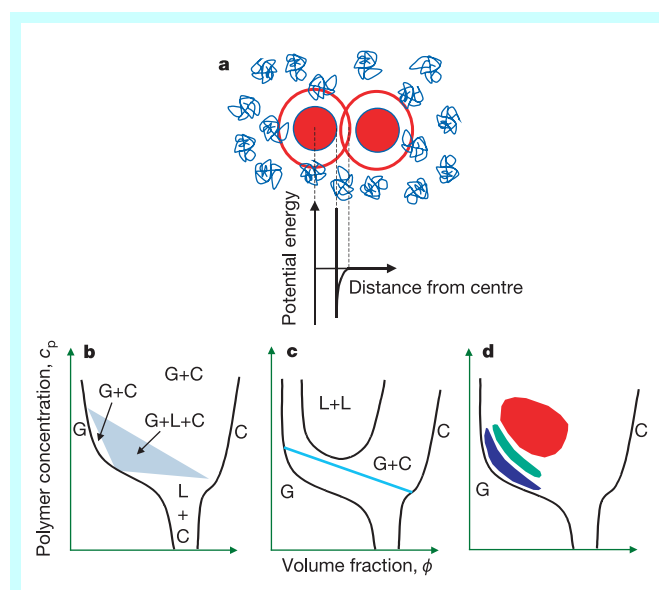


Figure 4 The depletion interaction and its phase diagrams, a concept that was introduced by Asakura and Oosawa⁷⁵ in 1954 and later by Vrij⁷⁶. **a**, A volume exists around each sphere into which the centre of the polymer coil cannot penetrate. On approach these 'excluded volumes' overlap and the total volume available to the polymers increases, increasing their entropy (at the expense of some colloidal entropy). Statistical mechanics predict the partitioning into colloid-rich and polymer-rich phases. For a given colloid size the phase diagrams are sensitive to the polymer size: **b**, large polymer; **c**, small polymer. The polymer can partition into different phases, so the lines indicating coexisting phases are not necessarily horizontal. Phase diagrams are similar to those in Fig. 1, but are inverted, as polymer concentration plays the role of inverse temperature. **b** is similar to an atomic system, having regions of three-phase equilibria (shaded area). **c** is similar to systems found for some proteins, with two-phase coexistence and a metastable liquid–liquid region. **d**, The situation often found experimentally, for short polymers. The blue area indicates a region where crystals are formed; the green area indicates aggregation; and the red area indicates gelation. Thus, only in the narrow region close to the phase boundary are the equilibrium phases reached.

Predictions of the kinetics often come from free-energy diagrams, from which the unstable and metastable regions and the commonly associated mechanisms of spinodal decomposition and nucleation and growth can be delineated. In spinodal decomposition, almost any fluctuation in the order parameter (here, colloid volume fraction) lowers the free energy and is commonly modelled by the Cahn–Hilliard theory using a coarse-graining approach (that is, considering length scales large compared to the colloid) to examine the change in volume fraction. However, the possibilities are more numerous than this. In the 1960s Cahn²⁴ described how free energy curves can predict which nucleation events are possible. In an extension of this work, Evans *et al.*²⁵ show that for a given position in the phase diagram several mechanisms may occur simultaneously and, equally important, some mechanisms may not be possible. Both gas nucleation and crystal nucleation are seen to occur concurrently²⁶; the kinetics in that part of the phase diagram will be very different to those where only crystal nucleation is feasible. Although examples have focused on work on colloid–polymer mixtures²⁷, the method by which possible mechanisms are identified depends only on the free-energy curves and is equally applicable to atomic, colloidal or protein systems.

Evans *et al.*²⁵ predict that for a small region of the phase diagram, the only mechanism of phase separation is by the nucleation of the solid phase (kinetic region E, in their terminology). Protein crystallographers have long been aware that the structures obtained depend sensitively on the initial conditions and this region, close to the phase boundary, seems to be that often giving the best crystals in both colloid and protein crystallization. Globular proteins and colloids have similar global phase diagrams but there are several differences. Proteins are neither perfectly spherical, nor isotropic. Surface groups often result in proteins having orientational order in the crystal and this anisotropic (orientation-dependent) interaction is likely to result in more complex mechanisms²⁸, while leaving the global phase diagram relatively unchanged.

Aggregation behaviour

The narrow window of crystallization (see Fig. 4d) mentioned above seems, at first, counter-intuitive. The drive to phase-separate increases on going further from the phase boundary, so we might assume that the rate of separation and crystal formation would also increase. In fact, although change is fast in such regions, the mechanism of change itself obstructs the pathway to crystallization and often crystals are not formed—a good demonstration of the proverb 'more haste, less speed'.

Different mechanisms can indeed be seen as the interaction

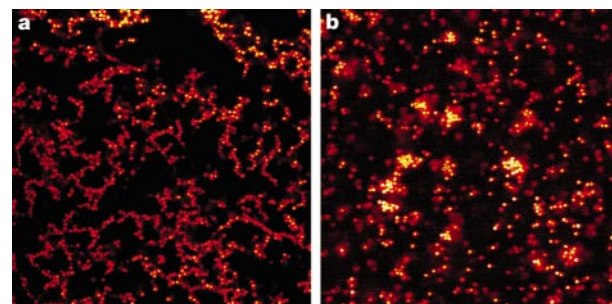


Figure 5 Confocal microscopy imaging at phase separation in colloid–polymer mixtures³⁵. Different pathways to phase separation are observed in different parts of the phase diagram. **a**, Open, stringy structures are observed deep within the phase-separation region where the dominant mechanism is aggregation. **b**, Closer to the binodal, the mechanism appears more like that of nucleation and growth: compact clusters appear after a significant delay time.

between particles increases (Fig. 5). Deep in the phase diagram, aggregation occurs. This is best described on a particle level. The first model was formulated at the beginning of the twentieth century by Schmoluchowski²⁹. Making the assumption that particles 'stick' immediately and permanently on contact, he successfully described the time evolution of the particle size distribution.

Since then there has been a large volume of work on aggregation using a number of different assumptions^{30–36}. Those above describe an extreme of aggregation: evolution is limited by the time taken for particles or clusters to diffuse and encounter one another and is thus termed diffusion-limited cluster aggregation (DLCA). Experiments and simulations have revealed that this results in a monodisperse cluster-size distribution, the clusters having a fractal dimension, $d_f = 1.8$. An alternative situation is that particles do not stick immediately because of an energy barrier: here, the rate-limiting step is to overcome the barrier and this is termed reaction-limited cluster aggregation (RLCA). RLCA produces more compact clusters ($d_f = 2.1$) and a polydisperse cluster-size distribution.

Light-scattering experiments have revealed growth laws and surprising similarities with spinodal decomposition^{32,34,37–40}. Although spinodal decomposition has been seen in systems with a stable gas–liquid equilibrium⁴¹ and one could argue that aggregation and spinodal decomposition are similar in that there appears to be no barrier to phase separation, it is less easy to see how the coarse-graining model of Cahn and Hilliard applies to aggregating fractals. Thus the extent of similarity and its interpretation is of interest^{34,37}. Because of the fractal nature of the clusters, aggregation can result in a gel, where a cluster spans the sample. Simulations provide many predictions of the aggregate structure and intriguing insights into the effects of different interactions. Work by Tanaka and Araki⁴², for example, indicates that hydrodynamic interactions between particles may lead to more open, stringy structures.

Samples often show behaviour intermediate between RLCA and DLCA. If the assumption of sticking permanently is invalid, as it must be if the samples are to reach the equilibrium phases, further evolution is quite involved^{30,43–46}. Pure rearrangements (rotations of side-arms, for example), changes in aggregation mechanism and particle detachment may all occur, possibly concurrently. Thus the routes by which the equilibrium phase is formed are complex. In a system where the particles are likely to stick immediately they are, by the same token, less likely to detach again and surrounding particles caging in a central one may restrict particle movement⁴⁷. So for a particle to leave a site and move to one more favourable will take longer than in a system where the depth of attraction is less, but where the attraction is less the phase separation will probably have been dominated by a nucleation and growth mechanism which already favours crystal formation. Thus, despite the large driving force, the road to crystallization is hindered all the way and samples will take a long time to reach equilibrium.

Gelation is a frequent result of aggregation. Such gels are often quite weak and their rheology and evolution is of interest^{44,46,48–51}. Several groups have examined transient gels^{41,52–55}, which maintain a relatively robust structure for a certain time before collapsing. Dynamic light scattering indicates that the kinetic behaviour at the colloid–gel transition is remarkably similar to that at the colloid–glass transition^{50,51}. This suggests that the jamming observed in gelation in colloidal suspensions could be related to the jamming responsible for glass formation. In turn, this led to the proposal⁵⁶ that gelation and the glass transition may be manifestations of more general jamming transitions⁵⁷ where, depending on the system, the controlling factor may be temperature, applied stress, interaction energy or volume fraction.

Aggregation, in many guises, is also seen in protein systems. Intriguingly, these reveal there may be another mechanism of growth hitherto unseen in colloidal systems, that is, growth of protein crystals via attachment (to the crystal) of single proteins

(monomers) or intermediate dimers or oligomers. Whether or not this occurs is a subject of great debate^{58–61}, but the kinetics of the aggregation and crystallization would be greatly affected.

The effect of metastable states

It was shown in Fig. 4 that as the range of attraction is reduced the fluid–fluid transition becomes metastable. Indeed, in long-range, three-phase systems, as the range is decreased and the stable fluid–fluid region almost disappears, the structure of the equilibrium fluid phase appears to change⁶². As illustrated in Fig. 6 even in the stable region the proximity of a fluid–fluid critical point may have a significant effect on the dynamics owing to the appearance of critical fluctuations, which may enhance crystal nucleation^{63,64}. The nucleus can also be enveloped by a fluid layer—this coating would alter the surface tension and thus the nucleation and growth process⁶⁵.

Within the metastable region the Ostwald rule implies that fluid–fluid phase separation would precede the fluid–crystal transition. Whether crystals are best grown inside or outside the metastable region is a matter of debate^{65–68}. If fluid–fluid separation occurs first, nucleation may be enhanced but growth will be determined by the now-high volume fraction and could produce less-well-ordered crystals. Outside the region, even with a fluid coating, growth is still determined by the average density. The fluid–fluid phase separation is not seen in colloidal systems, rather, as described above, it is usually pre-empted by a gelled state^{69,70}. In contrast, in protein systems, which have similar global phase diagrams, the transition is observed^{71–74}. This implies that there are subtle differences affecting the phase-separation pathways, and long-range forces may be important⁷⁰.

Thus phase transitions present us with a variety of bewildering pathways to phase separation, and although many are beginning to be well understood it is perhaps in those routes which lead to the non-equilibrium states—glasses or gels—where future surprises may be found. □

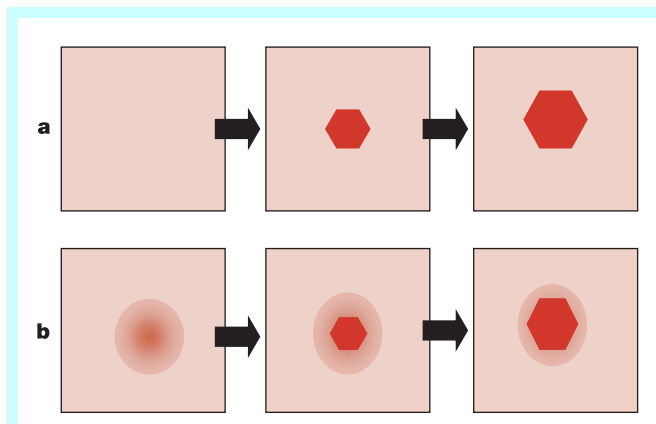


Figure 6 Nucleation mechanisms around the metastable critical point. **a**, The traditional picture of a crystal nucleating and growing in the original fluid. In order to get reasonable nucleation rates the initial concentration may have to be quite high, resulting in rapidly growing (owing to the large thermodynamic driving force) and thus more-disordered crystals, or aggregates. **b**, A two-step mechanism is suggested to occur around the metastable critical point. In this case, the initial step is the formation of a fluid droplet, either through phase separation⁶⁸ or the presence of critical fluctuations⁶³. The high density of this droplet increases the nucleation rate. As the crystal grows, a covering film of fluid acts like a buffer between the original fluid and the growing crystal. In protein systems this fluid layer aids growth because it means monomers are always available for the growth steps and they have ample time within the film to find the proper orientation for bonding. Thus the nucleation is increased because of the effect of metastable states on the dynamics, rather than by an increase in the thermodynamic driving force.

Received 19 July 2001; accepted 6 March 2002.

1. Auer, S. & Frenkel, D. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. *Nature* **409**, 1020–1023 (2001).
2. Elliot, M. S., Haddon, S. B. & Poon, W. C. K. Direct observation of pre-critical nuclei in a metastable hard-sphere fluid. *J. Phys. Condens. Matter* **13**, L553–L558 (2001).
3. Gasser, U., Weeks, E. R., Schofield, A., Pusey, P. N. & Weitz, D. A. Real space imaging of nucleation and growth in colloidal crystallization. *Science* **292**, 258–262 (2001).
4. Bruce, A. D., Wilding, N. B. & Ackland, G. J. Free energy of crystalline solids: a lattice-switch Monte Carlo method. *Phys. Rev. Lett.* **79**, 3002–3005 (1997).
5. Alexander, S. & McTague, J. P. Should all crystals be bcc? Landau theory of solidification and crystal nucleation. *Phys. Rev. Lett.* **41**, 702–705 (1978).
6. Zhu, J. *et al.* Crystallization of hard-sphere colloids in microgravity. *Nature* **387**, 883–885 (1997).
7. Pusey, P. N. *et al.* Structure of crystals of hard colloidal spheres. *Phys. Rev. Lett.* **63**, 2753–2756 (1989).
8. Pronk, S. & Frenkel, D. Can stacking faults in hard-sphere crystals anneal out spontaneously? *J. Chem. Phys.* **110**, 4589–4592 (1999).
9. Pusey, P. N. & van Megen, W. Phase-behaviour of concentrated suspensions of nearly hard colloidal spheres. *Nature* **320**, 340–342 (1986).
10. Debenedetti, P. G. & Stillinger, F. H. Supercooled liquids and the glass transition. *Nature* **410**, 259–267 (2001).
11. Angell, C. A., Ngai, K. L., McKenna, G. B., McMillan, P. F. & Martin, S. W. Relaxation in glassforming liquids and amorphous solids. *J. Appl. Phys.* **88**, 3113–3157 (2000).
12. Weeks, E. R., Crocker, J. C., Levitt, A. C., Schofield, A. & Weitz, D. A. Three dimensional direct imaging of structural relaxation near the colloidal glass transition. *Science* **287**, 627–631 (2000).
13. Dinsmore, A. D., Weeks, E. R., Prasad, V., Levitt, A. C. & Weitz, D. A. Three-dimensional confocal microscopy of colloids. *Appl. Opt.* **40**, 4152–4159 (2001).
14. Kegel, W. K. & van Blaaderen, A. Direct observation of dynamical heterogeneities in colloidal hard-sphere suspensions. *Science* **287**, 290–293 (2000).
15. Cui, B., Lin, B. & Rice, S. A. Dynamical heterogeneity in a dense quasi-two-dimensional colloidal liquid. *J. Chem. Phys.* **114**, 9142–9155 (2001).
16. Kegel, W. K. Crystallization in glassy suspensions of colloidal hard spheres. *Langmuir* **16**, 939–941 (2000).
17. Evans, R. M. L. & Holmes, C. B. Diffusive growth of polydisperse hard-sphere crystals. *Phys. Rev. E* **64**, 011404-1–011404-9 (2001).
18. Evans, R. M. L., Fairhurst, D. J. & Poon, W. C. K. Universal law of fractionation for slightly polydisperse systems. *Phys. Rev. Lett.* **81**, 1326–1329 (1998).
19. Bolhuis, P. G. & Kofke, D. A. Monte Carlo study of freezing of polydisperse hard spheres. *Phys. Rev. E* **54**, 634–643 (1996).
20. Kofke, D. A. & Bolhuis, P. G. Freezing of polydisperse hard spheres. *Phys. Rev. E* **59**, 618–622 (1999).
21. Auer, S. & Frenkel, D. Suppression of crystal nucleation in polydisperse colloids due to increase of the surface free energy. *Nature* **413**, 694–695 (2001).
22. Oxtoby, D. W. Crystallization: diversity suppresses growth. *Nature* **413**, 694–695 (2001).
23. Williams, S. R., Snook, I. K. & van Megen, W. Molecular dynamics study of the stability of the hard sphere glass. *Phys. Rev. E* **64**, 021506-1–021506-7 (2001).
24. Cahn, J. W. The metastable liquidus and its effect on the crystallization of glass. *J. Am. Ceram. Soc.* **52**, 118–121 (1969).
25. Evans, R. M. L., Poon, W. C. K. & Renth, F. Classification of ordering kinetics in three-phase systems. *Phys. Rev. E* **64**, 031403-1–031403-13 (2001).
26. Renth, F., Poon, W. C. K. & Evans, R. M. L. Phase transition kinetics in colloid-polymer mixtures at triple coexistence: kinetic maps from free-energy landscapes. *Phys. Rev. E* **64**, 031402-1–031402-9 (2001).
27. Poon, W. C. K. *et al.* Colloid-polymer mixtures at triple coexistence: kinetic maps from free-energy landscapes. *Phys. Rev. Lett.* **83**, 1239–1242 (1999).
28. Lomakin, A., Asherie, N. & Benedek, G. B. Aeolotropic interactions of globular proteins. *Proc. Natl Acad. Sci. USA* **96**, 9465–9468 (1999).
29. van Schmoluchowski, M. Versuch einer Mathematischen Theorie der Koagulationskinetik. *Z. Phys. Chem.* **92**, 129–168 (1917).
30. Asnaghi, D., Carpinetti, M. & Giglio, M. Recent developments in aggregation kinetics. *Mater. Res. Bull.* **19**(5), 14–18 (May 1994).
31. Poon, W. C. K. & Haw, M. D. Mesoscopic structure formation in colloidal aggregation and gelation. *Adv. Colloid Interface Sci.* **73**, 71–126 (1997).
32. Weitz, D. A., Huang, J. S., Lin, M. Y. & Sung, J. Limits of the fractal dimension for irreversible kinetic aggregation of gold colloids. *Phys. Rev. E* **54**, 1416–1419 (1985).
33. Asnaghi, D., Carpinetti, M., Giglio, M. & Sozzi, M. Coagulation kinetics and aggregate morphology in the intermediate regimes between diffusion-limited and reaction-limited cluster aggregation. *Phys. Rev. A* **45**, 1018–1023 (1992).
34. Poulin, P., Bibett, J. & Weitz, D. A. From colloidal aggregation to spinodal decomposition in sticky emulsions. *Eur. Phys. J. B* **7**, 277–281 (1999).
35. de Hoog, E. H. A., Kegel, W. K., van Blaaderen, A. & Lekkerkerker, H. N. W. Direct observation of crystallization and aggregation in a phase-separating colloid-polymer suspension. *Phys. Rev. E* **64**, 021407-1–021407-9 (2001).
36. Anderson, V. J., de Hoog, E. H. A. & Lekkerkerker, H. N. W. Mechanisms of phase separation and aggregation in colloid-polymer mixtures. *Phys. Rev. E* **65**, 011403-1–011403-8 (2002).
37. Hobbie, E. K. Depletion-driven phase separation and reversible aggregation in confined colloidal mixtures. *Langmuir* **15**, 8807–8812 (1999).
38. Rouw, P. W., Woutersen, A. T. J. M., Ackerson, B. J. & de Kruij, C. G. Adhesive hard sphere dispersions. *Physica A* **156**, 876–898 (1989).
39. Bibette, J., Mason, T. G., Gang, H. & Weitz, D. A. Kinetically induced ordering in gelation of emulsions. *Phys. Rev. Lett.* **69**, 981–984 (1992).
40. Carpinetti, M. & Giglio, M. Spinodal-type dynamics in fractal aggregation of colloidal clusters. *Phys. Rev. Lett.* **68**, 3327–3330 (1992).
41. Verhaegh, N. A. M., Asnaghi, D., Lekkerkerker, H. N. W., Giglio, M. & Cipolletti, L. Transient gelation by spinodal decomposition in colloid-polymer mixtures. *Physica A* **242**, 104–118 (1997).
42. Tanaka, H. & Araki, T. Simulation method of colloidal suspensions with hydrodynamic interactions: fluid particle dynamics. *Phys. Rev. Lett.* **85**, 1338–1341 (2000).
43. Hobbie, E. K. Metastability and depletion-driven aggregation. *Phys. Rev. Lett.* **81**, 3996–3999 (1998).
44. Bos, M. T. A. & van Opheusden, J. H. F. Brownian dynamics simulation of gelation and aging in interacting colloidal systems. *Phys. Rev. E* **53**, 5044–5050 (1996).
45. Shih, W. Y., Aksay, I. A. & Kikuchi, R. Reversible-growth model: cluster-cluster aggregation with finite binding energies. *Phys. Rev. A* **36**, 5015–5019 (1987).
46. Terao, T. & Nakayama, T. Sol-gel transition of reversible cluster-cluster aggregations. *Phys. Rev. E* **58**, 3490–3494 (1998).
47. Peters, E. A. J. F., Kollmann, M., Barenburg, T. M. A. O. & Philipse, A. P. Caging of a d-dimensional sphere and its relevance for the random dense sphere packing. *Phys. Rev. E* **63**, 21404-1–21404-9 (2001).
48. Gisler, T., Ball, R. C. & Weitz, D. A. Strain hardening of fractal colloidal gels. *Phys. Rev. Lett.* **80**, 778–781 (1998).
49. Krall, A. H. & Weitz, D. A. Internal dynamics and elasticity of fractal colloidal gels. *Phys. Rev. Lett.* **80**, 778–781 (1998).
50. Cipolletti, L., Manley, S., Ball, R. C. & Weitz, D. A. Glasslike kinetic arrest at the colloidal-gelation transition. *Phys. Rev. Lett.* **86**, 6042–6044 (2001).
51. Segre, P. N., Prasad, V., Schofield, A. B. & Weitz, D. A. Glasslike kinetic arrest at the colloidal-gelation transition. *Phys. Rev. Lett.* **86**, 6042–6044 (2001).
52. Verhaegh, N. A. M., Asnaghi, D. & Lekkerkerker, H. N. W. Transient gels in colloid-polymer mixtures studied with fluorescence confocal microscopy. *Physica A* **264**, 64–74 (1999).
53. Starrs, L., Poon, W. C. K., Hibberd, D. J. & Robins, M. M. Collapse of transient gels in colloid-polymer mixtures. *J. Phys. Condens. Matter* **14**, 2485–2505 (2002).
54. Evans, R. M. L. & Starrs, L. Emergence of a stress transmission lengthscale in transient gels. *J. Phys. Condens. Matter* **14**, 2507–2529 (2002).
55. Weeks, J. R., van Duineveldt, J. S. & Vincent, B. Formation and collapse of a gels of sterically stabilized colloidal particles. *J. Phys. Condens. Matter* **12**, 9599–9606 (2000).
56. Trappe, V., Prasad, V., Cipolletti, L., Segre, P. N. & Weitz, D. A. Jamming phase diagram for attractive particles. *Nature* **411**, 772–775 (2001).
57. Liu, A. J. & Nagel, S. R. Nonlinear dynamics: Jamming is not just cool any more. *Nature* **396**, 21–22 (1998).
58. Peters, R., Georgalis, Y. & Saenger, W. Accessing lysozyme nucleation with a novel dynamic light scattering detector. *Acta Crystallogr. D* **54**, 873–877 (1998).
59. Tanaka, S., Ito, K., Hayakawa, R. & Ataka, M. Size and number density of precrystalline aggregates in lysozyme crystallization process. *J. Chem. Phys.* **111**, 10330–10337 (1999).
60. Finet, S., Bonnet, E., Frouin, J., Provost, K. & Tardieu, A. Lysozyme crystal growth, as observed by small angle x-ray scattering, proceeds without crystallisation intermediates. *Eur. Biophys. J.* **27**, 263–271 (1998).
61. McPherson, A. *et al.* The effects of microgravity on protein crystallisation: evidence for concentration gradients around growing crystals. *J. Cryst. Growth* **196**, 572–586 (1999).
62. Moussaid, A., Poon, W. C. K., Pusey, P. N. & Soliva, M. F. Structure of marginal and fully developed colloidal liquids. *Phys. Rev. Lett.* **82**, 225–228 (1999).
63. Ten Wolde, P. R. & Frenkel, D. Enhancement of protein crystal nucleation by critical density fluctuations. *Science* **277**, 1975–1978 (1997).
64. Sear, R. Nucleation of a noncritical phase in a fluid near a critical point. *J. Chem. Phys.* **114**, 3170–3173 (2001).
65. Haas, C. & Drenth, J. The interface between a protein crystal and an aqueous solution and its effects on nucleation and crystal growth. *J. Phys. Chem. B* **104**, 368–377 (2000).
66. Poon, W. C. K. Crystallization of globular proteins. *Phys. Rev. E* **55**, 3762–3764 (1997).
67. Dixit, N. M. & Zukoski, C. F. Crystal nucleation rates for particles experiencing short-range attractions: Applications to proteins. *J. Colloid Interface Sci.* **228**, 359–371 (2000).
68. Haas, C. & Drenth, J. Understanding protein crystallization on the basis of the phase diagram. *J. Cryst. Growth* **196**, 388–394 (1999).
69. Soga, K. G., Melrose, J. R. & Ball, R. C. Metastable states and the kinetics of colloidal phase separation. *J. Chem. Phys.* **110**, 2280–2288 (1999).
70. Noro, M. G., Kern, N. & Frenkel, D. The role of long-range forces in the phase behaviour of colloids and proteins. *Europhys. Lett.* **48**, 332–338 (1999).
71. Thomson, J. A., Schurtenberger, P., Thurston, G. M. & Benedek, G. B. Binary liquid phase separation and critical phenomena in a protein/water solution. *Proc. Natl Acad. Sci. USA* **84**, 7079–7083 (1987).
72. Berland, C. R. *et al.* Solid-liquid phase boundaries of lens protein solutions. *Proc. Natl Acad. Sci. USA* **89**, 1214–1218 (1992).
73. Broide, M. L., Tominc, T. M. & Sazowsky, M. D. Using phase transitions to investigate the effect of salts on protein interactions. *Phys. Rev. E* **53**, 5325–5335 (1996).
74. Muschol, M. & Rosenberger, F. Liquid-liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization. *J. Chem. Phys.* **107**, 1953–1962 (1997).
75. Asakura, S. & Oosawa, F. On interaction between two bodies immersed in a solution of macromolecules. *J. Chem. Phys.* **22**, 1255–1256 (1954).
76. Vrij, A. Polymers at interfaces and the interaction in colloidal dispersions. *Pure Appl. Chem.* **48**, 471–483 (1976).

Acknowledgements

We thank D. Frenkel, E. H. A. de Hoog, W. K. Kegel and W. C. K. Poon for discussions about phase pathways in colloid-polymer and protein systems. This work was supported by the Council for Chemical Sciences from the Netherlands Organisation for Scientific Research (NOW-CW).

Correspondence and requests for materials should be addressed to H.N.W.L. (e-mail: h.n.w.lekkerkerker@chem.uu.nl).

The earliest known eutherian mammal

Qiang Ji*, Zhe-Xi Luo†, Chong-Xi Yuan*, John R. Wible‡, Jian-Ping Zhang‡ & Justin A. Georgi†

* Chinese Academy of Geological Sciences, Beijing 100037, China

† Carnegie Museum of Natural History, 4400 Forbes Avenue, Pittsburgh, Pennsylvania 15213, USA

‡ Geoscience University of China, Beijing 100083, China

The skeleton of a eutherian (placental) mammal has been discovered from the Lower Cretaceous Yixian Formation of northeastern China. We estimate its age to be about 125 million years (Myr), extending the date of the oldest eutherian records with skull and skeleton by about 40–50 Myr. Our analyses place the new fossil at the root of the eutherian tree and among the four other known Early Cretaceous eutherians, and suggest an earlier and greater diversification of stem eutherians that occurred well before the molecular estimate for the diversification of extant placental superorders (104–64 Myr). The new eutherian has limb and foot features that are known only from scansorial (climbing) and arboreal (tree-living) extant mammals, in contrast to the terrestrial or cursorial (running) features of other Cretaceous eutherians. This suggests that the earliest eutherian lineages developed different locomotory adaptations, facilitating their spread to diverse niches in the Cretaceous.

Placental mammals are the most diverse and dominant group of the three extant mammal lineages (Placentalia, Marsupialia and the egg-laying Monotremata)^{1,2}. All extant placentals, including our own order, Primates, are a subgroup of eutherians^{1,3}, which consist of extant placentals plus all extinct mammals that are more closely related to extant placentals (such as humans) than to extant marsupials (such as kangaroos). Here we describe a well-preserved eutherian mammal from the Early Cretaceous with bearing on the timing of the phylogenetic diversification and the locomotory evolution of the earliest eutherians.

Systematic palaeontology

Class Mammalia
Subclass Boreosphenida
Infraclass Eutheria
incertae sedis
Eomaia scansoria gen. et sp. nov.

Etymology. Eo (Greek): dawn; maia (Greek): mother; *Eomaia* for the earliest known eutherian mammal; *scansoria* (Latin) for the specialized skeletal features for climbing.

Holotype. CAGS01-IG1-a, b (Fig. 1; Chinese Academy of Geological Sciences, Institute of Geology), part and counterpart of a skeleton with an incomplete, flattened skull partially represented by impressions (dashed lines in Fig. 1), and nearly all of the post-cranium, with some preserved soft tissues, such as costal cartilages and fur.

Locality and horizon. Dawangzhangzi Locality, Lingyuan County, Liaoning Province, China. The holotype is from lacustrine silty shales of the Yixian Formation.

Geological age and fauna. The main fossiliferous horizon of the Yixian Formation was dated as 124.6 Myr (ref. 4) and correlated to the lower Barremian stage of the Mesozoic timescale⁵. The age of *Eomaia scansoria* is ~125 Myr and no younger than Mid Barremian. The Yixian Formation elsewhere in western Liaoning has yielded three other mammals, the spalacotheriid *Zhangheotherium*^{6,7}, the eutriconodont *Jeholodens*⁸, and the gobiconodontid *Repenomamus*⁹. The associated fauna in the Yixian Formation includes diverse fossil vertebrates, invertebrates and plants (reviewed in ref. 10).

Diagnosis. Among eutherians previously known from the late Early Cretaceous, *Eomaia scansoria* differs from *Prokennalestes*^{11,12} in lacking the labial mandibular foramen in the masseteric fossa and in having a larger metastylar and metaconal region on upper molar

M³; differs from *Murtoilestes*^{13,14} in having less-developed conules on the upper molars; and differs from both *Murtoilestes* and *Prokennalestes* in having an anteroposteriorly shorter molar trigonid and a longer talonid basin; differs from *Montanalestes*¹⁵ in having a paraconid lower than the metaconid; differs from the Late Cretaceous zhelestids^{16,17} and the Palaeocene ungulatiforms in lacking an inflated protocone and swollen lower molar cusps. *E. scansoria* differs from *Montanalestes*¹⁵ and all Late Cretaceous eutherians^{16–19} in retaining the primitive Meckel's sulcus on the mandible, and from most eutherians including placentals (but not *Prokennalestes*¹¹, *Montanalestes*¹⁵ and several asioryctitherians¹⁹) in having a slightly in-turned angular process of the mandible. *E. scansoria* differs from *Deltatheridium*^{20,21} and other metatherians²² (including marsupials) in having a typical eutherian dental formula, 5.1.5.3/4.1.5.3 (incisors, canine, premolars, molars) (Fig. 2); differs from most marsupials in lacking the hypoconulid shelf and having nearly equal distance between the talonid cusps^{20–23}; differs from stem boreosphenidans^{24,25} in possessing a larger entoconid of nearly equal size to the hypoconid; differs from nontribosphenic therians^{6,7,26,27} in having a tricuspid talonid in occlusion with protocone; differs from *Ausktribosphenos* and *Bishops* (eutherians by some^{28,29}, but endemic southern mammals by others^{24,25}) in lacking a shelf-like mesial cingulid on the molars, and in having laterally compressed ultimate and penultimate lower premolars without full cusp triangulation; differs from *Ausktribosphenos*, *Bishops* and most stem mammaliaforms in lacking the primitive postdentary trough on the mandible.

Description and comparison

Numerous skeletal apomorphies (evolutionarily derived characters) also distinguish *Eomaia* from currently known eutherians^{3,30–32}, the earliest known metatherians (including marsupials)^{22,23,33–35}, and nontribosphenic therians^{6,7,26,27}. The scapula is slender with a prominent coracoid process on the glenoid and a relatively large acromion process on a tall scapular spine. The clavicle is robust and curved, with its proximal end abutting the lateral process of the clover-shaped manubrium. *Eomaia* differs from the Late Cretaceous eutherians *Asioryctes* and *Zalambdalestes* in having an enlarged and elongate trapezium in the wrist (Fig. 3). The hamate is large, a feature shared probably by all trechnotherians, although not so large as the hypertrophied hamate of marsupials³⁴. The trapezoid and capitata are small, and their proportions to the hamate and trapezium are comparable to the condition in the grasping hands of living scansorial and arboreal mammals^{30,35}. *Eomaia* and other eutherians retain the primitive mammalian condition in which the

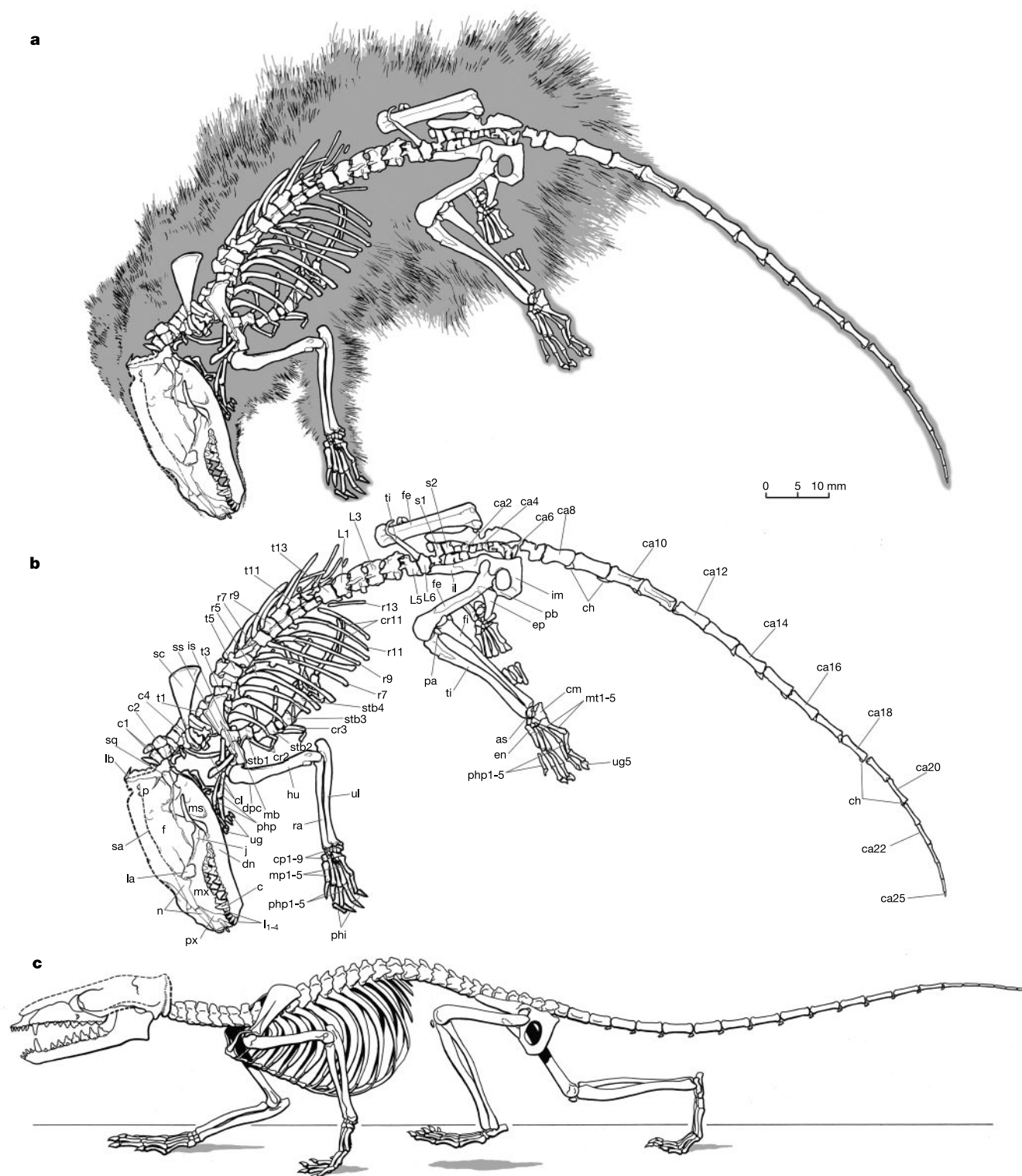


Figure 1 *Eomaia scansoria* (Chinese Academy of Geological Sciences (CAGS) 01-IG-1a, b; holotype). **a**, Fur halo preserved around the skeleton (01-IG-1a, many structures not represented on this slab are preserved on the counter-part 01-IG-1b, not illustrated). **b**, Identification of major skeletal structures of *Eomaia*. **c**, Reconstruction of *Eomaia* as an agile animal, capable of climbing on uneven substrates and branch walking. as, astragalus; c, canine; c1–c7, cervical vertebrae 1–7; ca1–ca25, caudal vertebrae 1–25; ch, chevron (caudal haemal arch); cl, clavicle; cm, calcaneum; cp1–9, carpals 1–9; cr1–11, costal cartilages 1–11; dn, dentary; dpc, deltopectoral crest; en, entocuneiform; ep, epipubis; f, frontal; fe, femur; fi, fibula; hu, humerus; l_{1–4}, lower incisors 1–4; il, ilium;

im, ischium; is, infraspinous fossa of scapula; j, jugal; la, lacrimal; lb, lambdoidal crest; L1–L6, lumbar vertebrae 1–6; mb, manubrium sterni; mp1–5, metacarpals 1–5; ms, masseteric fossa; mt1–mt5, metatarsals 1–5; mx, maxillary; n, nasal; p, parietal; pa, ossified patella; pb, pubis; phi, intermediate phalanges; php1–5, proximal phalanges 1–5; px, premaxillary; ra, radius; r1–r13, thoracic ribs 1–13; s1, s2, sacral vertebrae 1 and 2; sa, sagittal crest; sc, scapula; sq, squamosal; ss, supraspinous fossa of scapula; stb1–5, sternbrae 1–5 (sternbra 5 is the xiphoid); ti, tibia; t1–t13, thoracic vertebrae 1–13; ug1–5, ungual claws 1–5; ul, ulna.

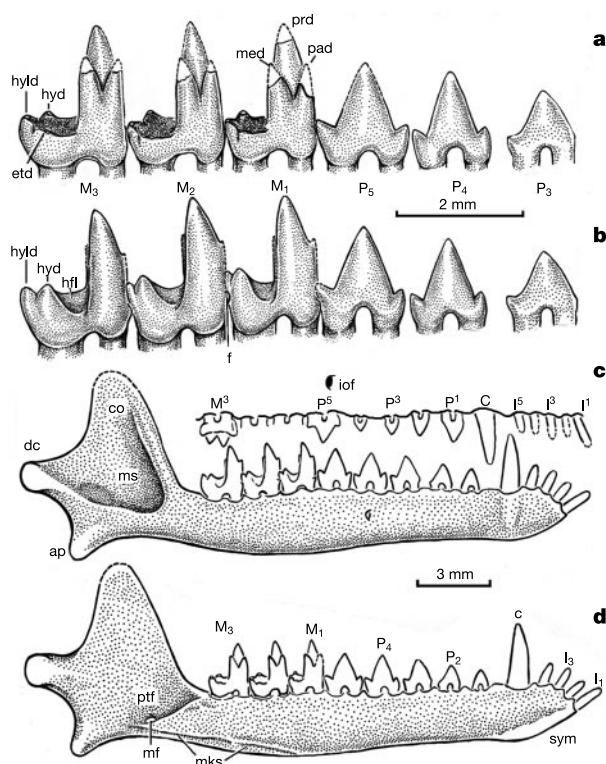


Figure 2 *Eomaia scansoria* dentition and mandible (composite reconstruction). **a**, Lower P₃–M₃ (right, lingual view). **b**, Lower P₃–M₃ (right, labial view). **c**, Upper dentition (incomplete) and mandible (right, lingual view). **d**, Mandible (left, lingual view). ap, angular process; C and c, upper and lower canine; co, coronoid process of mandible; dc, dentary condyle (articular process); etd, entoconid; f, cuspule f (anterolabial cingulid cuspule for interlocking); hfl, hypoflexid; hyd, hypoconid; hyd hypoconulid; I_{1–5} and I_{1–4}, upper and lower incisors; M_{1–3} and M_{1–3}, upper and lower molars; med, metaconid; mf, posterior (internal) foramen of mandibular canal; mks, Meckel's sulcus; ms, masseteric fossa; P_{1–5} and P_{1–5}, upper and lower premolars; pad, paraconid; prd, protoconid; ptf, pterygoid muscle fossa; sym, mandibular symphysis.

scaphoid and the triquetrum are small relative to other carpals, in contrast to the hypertrophied scaphoid and triquetrum in marsupials^{34,35}. Metacarpal 5 is level with the anterior edge of the hamate, a synapomorphy of crown therians, in contrast to the primitive *Zhangheotherium*⁷ and *Jeholodens*⁸, in which the proximal end of metacarpal 5 overhangs and is offset from the hamate (Fig. 3).

The ilium, ischium and pubis are fused. The epipubis is present. Relative to the size of the pelvis and the obturator foramen, the pubic symphysis is much shorter than those of the nontribosphenic therians *Zhangheotherium*, *Henkelotherium*²⁶ and *Vincelestes*²⁷, and early Tertiary metatherians²³. Given the short sacral transverse processes and the deep pelvis, it is likely that the pelvis was narrow at the sacral joint and vertically deep, as in *Zalambdalestes*³¹, *Ukhaatherium*³ and multituberculates^{36,37}, but less like the shallow pelvis of early Tertiary metatherians²³. The patella is present on both hindlimbs, a derived placental feature that is absent in most of the basal metatherians³⁴, and absent in all known skeletons of nontribosphenic trechnotherians, but convergent to multituberculates^{36–38} and monotremes³⁸.

The ankle of *Eomaia* bears strong resemblance to those of Late Cretaceous eutherians^{30–32}. The medial astragalotibial facet is well developed, vertical, and separated by a sharp crest from the lateral astragalotibial facet, a diagnostic eutherian feature. The navicular facet is distinctive from the astragalar neck, and does not extend to the medial side of the neck as in metatherians^{34,35}. The entocuneiform is elongate, and its joint with metatarsal 1 is offset anteriorly

from the joint of the intermediate cuneiform and metatarsal 2 (Fig. 4). The calcaneum is similar to those of *Asioryctes*³⁰, *Ukhaatherium*³² and *Zalambdalestes*^{31,32} in retaining many primitive features of the crown therians^{23,32,34}. The absence of the fibular malleolus is a phylogenetically primitive character, but nonetheless permitting a partial eversion of the foot³⁷.

Hairs are preserved as carbonized filaments and impressions around most of the body, although their traces are thin on the tail (Fig. 1a). The pelage appears to have both guard hairs and a denser layer of underhairs close to the body surface. Fossilized hairs were previously reported in Tertiary placentals and multituberculate mammals³⁹. *Eomaia* adds to the evidence that presence of hairs is a ubiquitous feature of mammals.

Scansorial adaptation

The fore- and hindfeet of *Eomaia* (Figs 4 and 5) show similar phalangeal proportions and curvature to the grasping feet of extant arboreal mammals, such as the didelphid *Caluromys*³⁵, the flying lemur *Cynocephalus*, and arboreal primates^{40,41}. In phalangeal features, *Eomaia* is more similar to arboreal mammals than to such scansorial taxa as the tree shrew and opossum. The proximal manual phalanx is arched dorsally (Fig. 5). Its proximal, ventral surface has a shallow longitudinal groove for the tendon of flexor digitorum profundus. Two protuberances for the fibrous tendon sheaths of the flexor digitorum flank the sides of the phalanx three-quarters of the length towards the distal end, which is partially trochleated. Sesamoid bones are present at both interphalangeal joints. *Eomaia* is similar to the scansorial or fully arboreal archontan eutherians^{40,41} in all these characteristics of the proximal phalanges (Fig. 5). The proportions of the intermediate phalanx to the proximal phalanx differ among terrestrial, scansorial and fully arboreal didelphid marsupials³⁵ and placental carnivores. This ratio in *Eomaia* (Fig. 5f) is intermediate between the fully arboreal *Micoureus* and *Caluromys*, and the scansorial *Didelphis* and the fully terrestrial *Metachirus*³⁵. Pedal digits 4 and 5 are elongate relative to the medial digits (1, 2 and 3). Both proximal and intermediate phalanges of digits 4 and 5 (although not the metatarsals) are longer than their counterparts in digit 3 (Fig. 4), a convergent pattern of many unrelated scansorial and arboreal mammals³⁶, in contrast to terrestrial or cursorial mammals, in which the phalanges in digits 2 and 3 are longer than those in digits 4 and 5 (ref. 36).

Both manual and pedal claws are more similar to those of scansorial mammals^{36,42}, than to fully arboreal taxa⁴². The pedal claws have an arched dorsal margin, a large flexor tubercle, and a small dorsal lip on the proximal articular surface for the extensor insertion. These are identical in lateral profile to those of the dormouse *Glis*, an extant rodent active in low bushes³⁶, and consistent with the claw morphotype of all extant scansorial mammals⁴². The preserved impression of the manual claw lacks the broad and thickened dorsal margin found in *Jeholodens*⁸ and *Zhangheotherium*⁷, so we interpret the claw to be more laterally compressed beyond the articulating margin (Fig. 3) than in the latter taxa. These are typical features of scansorial mammals⁴², such as the tree shrew *Tupaia*^{40,41} and the rodent *Glis glis*³⁶.

Other features of *Eomaia* are also consistent with an arboreal or scansorial adaptation: well-developed scapular acromion and coracoid process plus a tall spine^{23,35}, caudal vertebral column twice the length of the precaudal vertebral column^{26,37}, and elongation of the mid-caudal vertebrae²⁶. These convergences strongly suggest that *Eomaia* was an agile animal with climbing skeletal adaptations, capable of grasping and branch walking, and active both on the ground and in trees or shrubs (for example, like the opossum *Didelphis*, some species of the tree shrew *Tupaia*, and the dormouse *Glis*). We estimate that the holotype of *Eomaia scansoria* had a body mass from 20 to 25 g. For such small mammals, some capacity for climbing is required for moving on uneven substrates even in a terrestrial habitat⁴³; therefore, the anatomical differences between

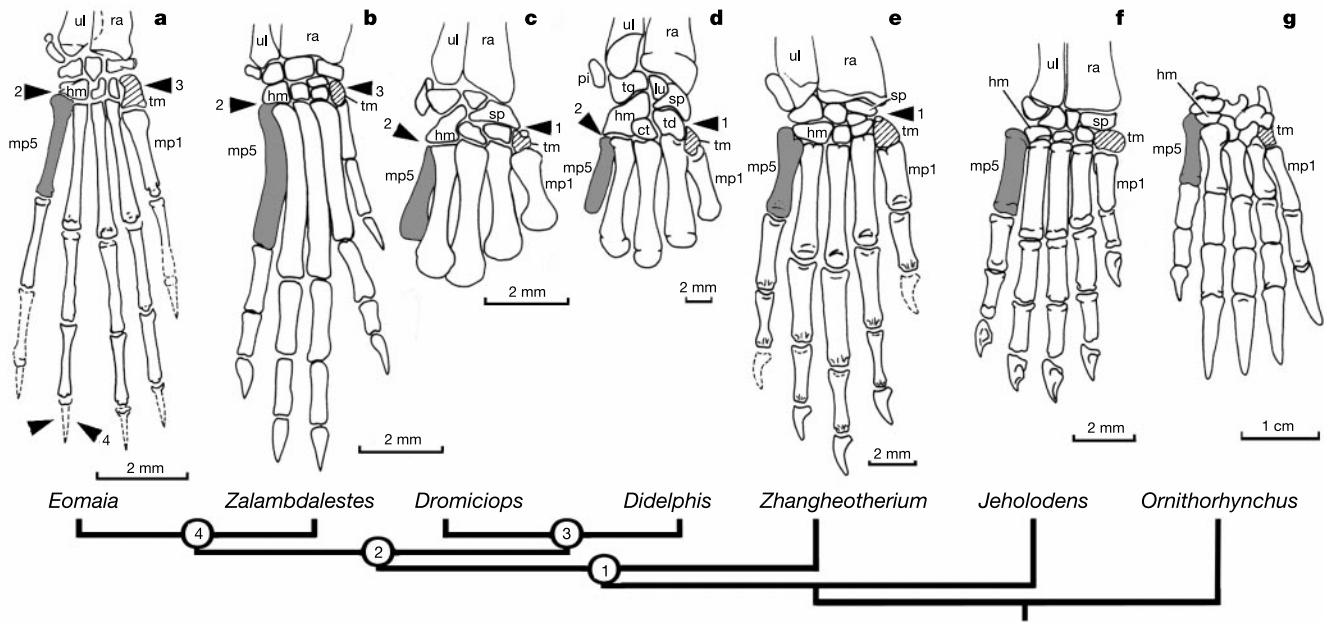


Figure 3 Comparison of forefoot (manus) of *Eomaia scansoria* (dorsal view, right). **a**, Eutherian *Eomaia* (composite reconstruction). **b**, Eutherian *Zalambdalestes*³¹. **c**, Marsupial *Dromiciops*³⁴. **d**, Marsupial *Didelphis*^{30,34}. **e**, Trechnotherian *Zhangheotherium*⁷. **f**, Eutriconodont *Jeholodens*⁸. **g**, Monotreme *Ornithorhynchus*. Apomorphies are as follows. Node 1 (Trechnotheria): enlargement of hamate, elongation of metacarpals, trapezium (hatched) offset from scaphoid (arrow 1; but reversed in eutherians). Node 2 (crown Theria): longitudinal alignment of mp5 (shaded) to hamate

(arrow 2; in contrast to the plesiomorphy of mp5 being offset from hamate; but reversed in some placentals), presence of distal radial malleolus. Node 3 (Marsupialia): hypertrophy of hamate, enlargement of scaphoid and triquetrum³⁴. Node 4 (Eutheria): elongation (longer than wide) of trapezium (hatched) (arrow 3; in contrast to the primitive condition of being wider than long). *Eomaia*: laterally compressed claws (arrow 4). ct, capitata; hm, hamate; lu, lunate; mp1–5, metacarpals 1–5; pi, pisiform; ra, radius; sp, scaphoid; td, trapezoid; tm, trapezium; tq, triquetrum; ul, ulna.

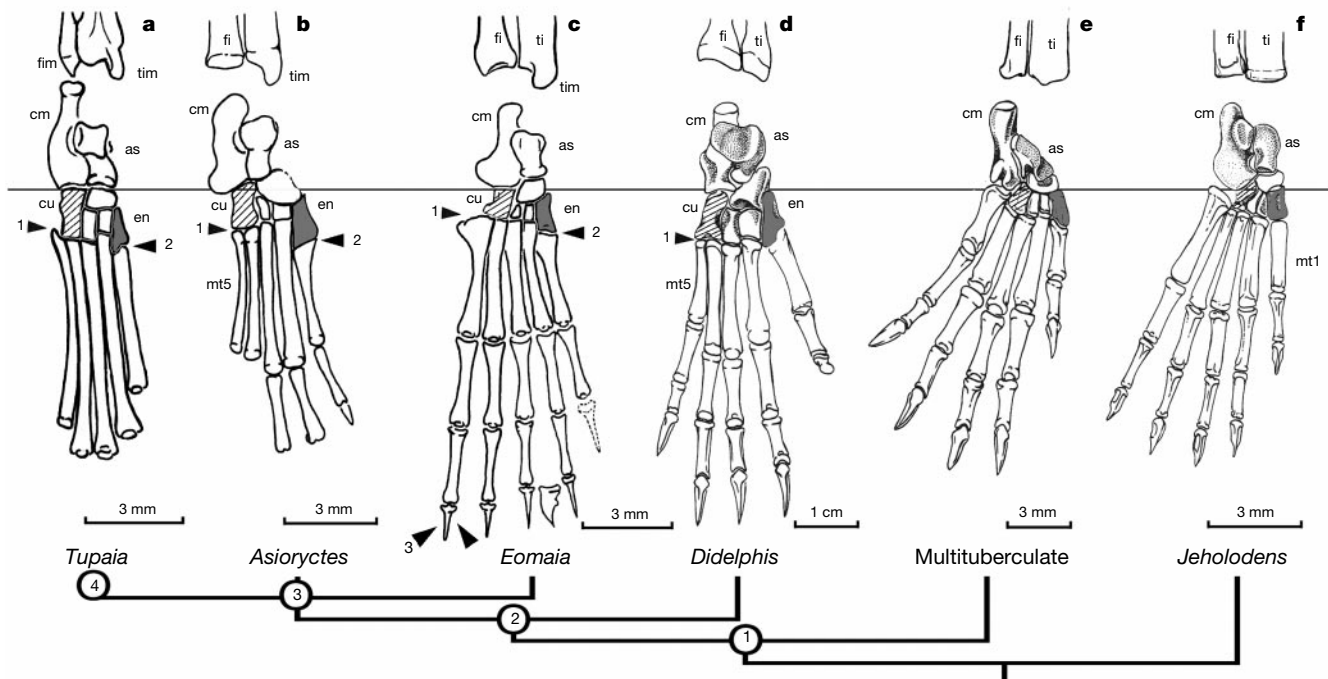


Figure 4 Comparison of hindfoot (pes) of *Eomaia scansoria*. **a**, Placental *Tupaia*³⁰. **b**, Eutherian *Asioryctes*³⁰. **c**, Eutherian *Eomaia* (composite reconstruction). **d**, Marsupial *Didelphis*. **e**, Multituberculata³⁶. **f**, Eutriconodont *Jeholodens*. Apomorphies are as follows. Node 1 (Therianformes³⁹): partial superposition of astragalus and calcaneum, laterally compressed calcaneal tuber. Node 2 (crown Theria): elongation and enlargement of cuboid; distal alignment of cuboid and metatarsal 5 so that the cuboid (hatched) corresponds to both metatarsals 4 and 5 (arrow 1; in contrast to the primitive condition of

mt5 being offset from cuboid). Node 3 (Eutheria): enlarged tibial malleolus; mt1–entocuneiform (shaded) joint offset from the mt2–mesocuneiform joint (arrow 2; in contrast to the primitive condition of the mt1–entocuneiform joint level with the mt2–mesocuneiform joint). Node 4 (Placentalia): fibular malleolus and the complete mortise-tenon upper ankle joint^{32,34}. *Eomaia*: compressed ungual claw (arrow 3). as, astragalus; cm, calcaneum; cu, cuboid; en, entocuneiform; fi, fibula; fim, fibular distal malleolus; mt1–5, metatarsals 1–5; ti, tibia; tim, tibial distal malleolus.

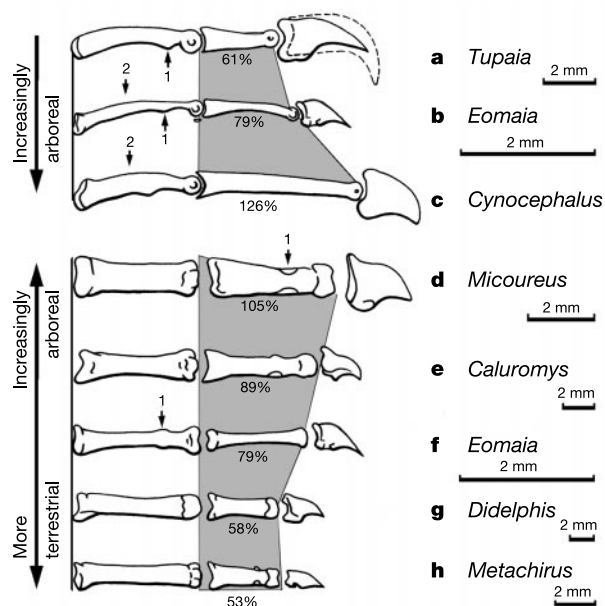


Figure 5 Comparison of manual phalanges and claws of *Eomaia scansoria* to those of scansorial and arboreal placentals (lateral view of digit 3). **a** Tree shrew *Tupaia* (scansorial, after refs 40 and 41). **b**, *Eomaia* (based on CAGS01-IG-1b). **c**, Flying lemur *Cynocephalus* (arboreal, after ref. 40). Comparison to manual phalanges of didelphid marsupials (digit 3; proximal and intermediate phalanges in ventral view; claw in lateral view). **d**, *Micoureus* (fully arboreal, after ref. 35). **e**, *Caluromys* (fully arboreal, after ref. 35). **f**, *Eomaia*. **g**, Opossum *Didelphis* (scansorial, after ref. 35). **h**, *Metachirus* (fully terrestrial, after ref. 35). The proximal phalanges standardized to the same length; percentage represents the length ratio of the intermediate to the proximal phalanges; scale varies among taxa. Arrow 1, protuberance on phalanges for the fibrous tendon sheaths for the flexor digitorum profundus (on the proximal phalanx in eutherians, but on the intermediate phalanx in didelphid marsupials). Arrow 2, dorsal curvature typical of scansorial or arboreal eutherians.

scansoriality and arboreality are not significant^{43,44}. The available evidence is insufficient to determine whether *Eomaia* was scansorial (as in some species of *Tupaia* or *Glis*) or fully arboreal (for example, as the marsupial *Caluromys* and the tupaiid *Ptilocercus*).

Because most basal metatherians are scansorial^{23,34,35}, scansorial skeletal features appear to be primitive for the earliest known eutherians. But the evidence for an ancestral scansorial adaptation for the crown group therians as a whole is less clear, because the successive outgroups to crown therians are either scansorial (for example, *Henkelotherium*²⁶) or terrestrial (for example, *Vincelestes*²⁷ and *Zhangheotherium*⁷).

Phylogenetic relationships

On the basis of 268 characters sampled from all major Mesozoic mammal clades and principal eutherian families of the Cretaceous, *Eomaia* is placed at the root of the eutherian tree with *Murtoilestes* and *Prokennalestes*. Clearly, these three taxa are closer to living placentals than to living marsupials. *Eomaia* is placed in Eutheria (Fig. 6) by numerous apomorphies in the dentition (Fig. 2), the wrist (Fig. 3) and the ankle (Fig. 4). Among eutherians, *Eomaia* is similar to *Prokennalestes* because they have identical features in lower premolars P₄ and P₅^{11,12}, plus identical and unique features of the posterior mandibular foramen on the ventral margin of the pterygoid fossa, with Meckel's sulcus intersecting the margin of the pterygoid fossa posterior to the mandibular foramen (Fig. 2). *Murtoilestes*^{13,14} is similar to *Prokennalestes* in molar characteristics.

Our analysis, including information of *Eomaia*, corroborates several previous hypotheses of relationships among the Late Cretaceous eutherians (Fig. 6a). First, asioryctitherians³ from the Campa-

nian (~75 Myr) of Mongolia form a clade^{3,21} that also includes zambdalestids. Second, zhelestids from the Coniacian (>85 Myr) of Uzbekistan may be related to ungulatormorphs from the Tertiary of North America^{16,17}. Third, the ungulatormorphs, in a successively more distant order, may be related to the palaeoryctid *Cimolestes* and the leptictid *Gypsonictops* from the Maastrichtian (~70 Myr) of North America, and possibly to the Coniacian eutherian *Daulestes*. Fourth, the North American *Montanalestes*¹⁵ is placed among the basal eutherians, although its position differs in ordered and unordered searches (see Supplementary Information). Our analysis did not include enough extant placental superorders to address

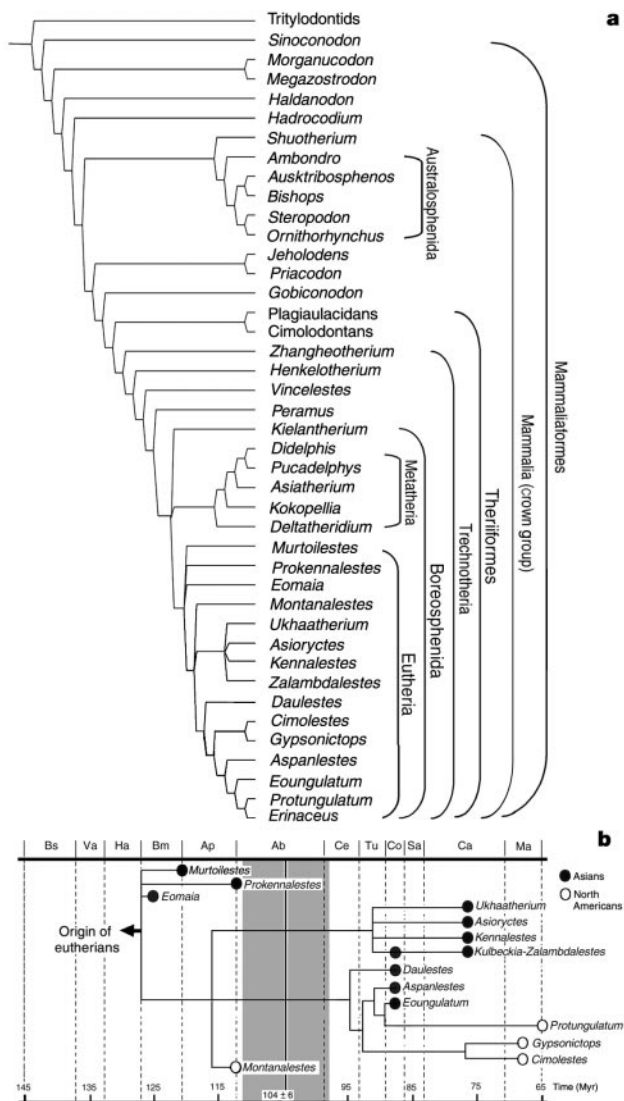


Figure 6 Phylogeny of eutherian *Eomaia scansoria* (**a**) and timing of the earliest evolution of eutherians (**b**). The phylogeny of mammals is based on a strict consensus of 50 equally parsimonious trees (tree length = 919, consistency index = 0.508, retention index = 0.740) from a PAUP analysis (version 4.0b, 1,000 runs of heuristic search, with unordered multistate characters) of 268 dental and skeletal characters that can be scored for the comparative taxa (the topology from searches with some ordered multistate characters is presented in Supplementary Information). The minimal age of *Eomaia* is after ref. 4; the age estimate for *Murtoilestes* is after ref. 13; dating of the Uzbekistan eutherians is after refs 16 and 17; and dating of the Mongolian eutherians is after ref. 45. The earliest molecular estimate^{46,47} of divergence of superordinal placental clades is 104 Myr (also see refs 48 and 49). Cretaceous stages: Ab, Albian; Ap, Aptian; Bm, Barremian; Bs, Berriasian; Ca, Campanian; Ce, Cenomanian; Co, Coniacian; Ha, Hauterivian; Ma, Maastrichtian; Sa, Santonian; Tu, Turonian; Va, Valanginian.

whether some Cretaceous eutherians could be linked to placental superordinal clades, as advocated by some^{16,17}, or these eutherians are extinct lineages unrelated to living placentals, as argued by others^{3,21}.

Earliest eutherian diversification

Because *Eomaia*, *Prokennalestes* and *Murtoilestes* are distinct from each other, and from the three other previously recognized clades of *Montanalestes*, asioryctitherians–zalambdalestids, and zhelestids–ungulatomorphs (Fig. 6), the diversification of some of these clades must have occurred by the first appearance of *Eomaia* (~125 Myr). The second-oldest eutherian, *Murtoilestes*, is near the Barremian–Aptian transition¹³, also consistent with the early diversification of these taxa in the Barremian (~127–121 Myr). *Montanalestes* from the Aptian–Albian of North America is contemporary with (but distinctive from) all Barremian–Albian taxa of Asia, from which it must have split before Aptian–Albian. The ages of these taxa and their basal positions in the eutherian family tree (Fig. 6a) suggest the divergence of such typically Late Cretaceous lineages as asioryctitherians–zalambdalestids and zhelestids–ungulatomorphs (or the divergence of a combination of these clades) was no later than Aptian–Albian (Fig. 6a, tree topology from unordered searches), and could be as early as Barremian (tree topology from ordered searches, see Supplementary Information). Concurrent with their phyletic splitting, these lineages also developed different locomotory specializations that may have facilitated their spread to different niches. *Eomaia* was a scansorial animal in a lake-shore environment^{6–10}. In contrast to *Eomaia*, most (but not all) asioryctitherians were terrestrial^{30,32} in a xeric sand-dune niche⁴⁵. Most ungulatomorphs (including zhelestids), palaeoryctids and leptictids were in a fluvial palaeoenvironment¹⁶, although zalambdalestids with saltatorial locomotion³¹ are known from both the dune-field⁴⁵ and fluvial sediments¹⁶.

Given the earliest eutherian phylogeny (Fig. 6), the dating of *Eomaia* and *Murtoilestes* provides the minimal time for the diversification of stem eutherians at 125 Myr. The molecular dating done most recently shows that the placental superordinal diversification has ranged from 64 to 104 Myr^{46,47}. The earliest minimal quartet estimate (the split of xenarthrans and cetartiodactyls) is 104 Myr (95% confidence interval ± 6 Myr)^{46,47}. The divergence time of the molecular superordinal clades I (afrotheres) and II (xenarthrans) from III and IV (other placentals) ranges from ~72 to ~112 Myr^{46,47} (at 95% confidence intervals). Although some previous estimates⁴⁸ postulated a 129 ± 18.5 Myr split for xenarthrans, the statistically defensible date for the split of the superordinal placental clades is now reconsidered by the same authors⁴⁹ to be about 112 ± 7 Myr (based on the hystricognaths–sciurognaths split). The minimal divergence time (≥ 125 Myr) of the stem eutherian lineages on the basis of the most recently discovered fossils is consistent with either of the two molecular estimates for divergence of the placental superordinal clades (104 ± 6 Myr^{46,47} or 112 ± 7 Myr by revision⁴⁹ of ref. 48). The time sequence of eutherian diversification (dated by fossils) and the splitting of the extant placental superorders (dated by the molecular clock) are consistent with the phylogenetic sequence in which multiple eutherian stem clades had split (Fig. 6b) well before the radiation of the extant placental superordinal clades. The corroboration of fossils and molecules provides a timetable of the earliest eutherian–placental evolution for calibrating the rates of morphological^{1,2} and molecular^{46–49} evolution.

Previous molecular studies postulated a gap between the putative time of origin of the placental superordinal groups and the then earliest fossil record of stem eutherians (for example, ref. 48). This gap was so large that it conflicted with the preservation likelihood models for eutherians based on empirical assessment of the mammalian fossil records⁵⁰. Discoveries of *Eomaia* and *Murtoilestes*, and the upward revision of molecular dating (for example, refs 46 and 47), have eliminated this putative gap. Whether, or to what extent,

the revised earliest eutherian records (≥ 125 Myr) would be consistent with the recently revised molecular estimates (for example, refs 46–49) can be further tested by the preservation likelihood models⁵⁰. □

Methods

The holotype of *Eomaia scansoria* suffered light diagenetic metamorphism, common for shales of the Yixian Formation. Some bones and all teeth are frail and fractured. However, the impressions of these structures are preserved in excellent detail. Composite restorations of the dentition, manus and pes are based on the outlines by camera lucida tracing; topographic features are based on reversed stereophotos from digital images of the well-preserved impressions. Relationships of *Eomaia* (Fig. 6a) are based on parsimony analysis of all major clades of Mesozoic mammals including the principal families of Cretaceous eutherians, plus the southern tribosphenic mammal *Ausktribosphenos*, considered by some to be eutherian^{28,29}. *Erinaceus* is included here to show that *Ausktribosphenos* and *Bishops* are not related to erinaceids (after refs 24 and 25). A total of 268 dental and skeletal characters was sampled for this analysis. These characters are combined from several recent analyses to resolve relationships of mammaliaforms, southern tribosphenic mammals, eutriconodonts, multituberculates, trechnotherians, stem boreosphenidans, metatherians and eutherians (sources are listed in Supplementary Information).

Received 14 November 2001; accepted 15 March 2002.

- McKenna, M. C. & Bell, S. K. *Classification of Mammals above the Species Level* 1–631 (Columbia Univ. Press, New York, 1997).
- Novacek, M. J. Mammalian phylogeny—shaking the tree. *Nature* **356**, 121–125 (1992).
- Novacek, M. J. *et al.* Epipubic bones in eutherian mammals from the Late Cretaceous of Mongolia. *Nature* **389**, 483–486 (1997).
- Swisher, C. C. III, Wang, Y.-Q., Wang, X.-L., Xu, X. & Wang, Y. Cretaceous age for the feathered dinosaurs of Liaoning, China. *Nature* **398**, 58–61 (1999).
- Gradstein, F. M. *et al.* in *Geochronology, Time Scales and Global Stratigraphic Correlation* (eds Berggren, W. A., Kent, D. V., Aubry, M. P. & Hardenbol, J.) 95–126 (Spec. Pub. No. 54, SEPM, Soc. Sed. Geol., Tulsa, Oklahoma, 1995).
- Hu, Y.-M., Wang, Y.-Q., Luo, Z.-X. & Li, C.-K. A new symmetrodont mammal from China and its implications for mammalian evolution. *Nature* **390**, 137–142 (1997).
- Hu, Y.-M., Wang, Y.-Q., Li, C.-K. & Luo, Z.-X. Morphology of dentition and forelimb of *Zhangheotherium*. *Vert. Palasiatic* **38**, 102–125 (1998).
- Ji, Q., Luo, Z.-X. & Ji, S.-A. A Chinese triconodont mammal and mosaic evolution of the mammalian skeleton. *Nature* **398**, 326–330 (1999).
- Wang, Y.-Q., Hu, Y.-M., Meng, J. & Li, C.-K. An ossified Meckel's cartilage in two Cretaceous mammals and origin of the mammalian middle ear. *Science* **294**, 357–361 (2001).
- Gee, H. (ed.) *Rise of the Dragon—Readings from Nature about the Chinese Fossil Records* 1–262 (Univ. Chicago Press, Chicago, 2001).
- Kielan-Jaworowska, Z. & Dashzeveg, D. Eutherian mammals from the Early Cretaceous of Mongolia. *Zool. Scripta* **18**, 347–355 (1989).
- Sigogneau-Russell, D., Dashzeveg, D. & Russell, D. E. Further data on *Prokennalestes* (Mammalia Eutheria inc. sed.) from the Early Cretaceous of Mongolia. *Zool. Scripta* **21**, 205–209 (1992).
- Averianov, A. O. & Skutschas, P. P. A eutherian mammal from the Early Cretaceous of Russia and biostratigraphy of the Asian Early Cretaceous vertebrate assemblages. *Lethaia* **33**, 330–340 (2000).
- Averianov, A. O. & Skutschas, P. P. A new genus of eutherian mammal from the Early Cretaceous of Transbaikalia, Russia. *Acta Palaeontol. Pol.* **46**, 431–436 (2001).
- Cifelli, R. L. Tribosphenic mammal from the North American Early Cretaceous. *Nature* **401**, 363–366 (1999).
- Nessov, L. A., Archibald, J. D. & Kielan-Jaworowska, Z. Ungulate-like mammals from the Late Cretaceous of Uzbekistan and a phylogenetic analysis of Ungulatomorpha. *Bull. Carnegie Mus. Nat. Hist.* **34**, 40–88 (1998).
- Archibald, J. D., Averianov, A. O. & Ekdale, E. G. Late Cretaceous relatives of rabbits, rodents, and other placental mammals. *Nature* **414**, 62–65 (2001).
- Kielan-Jaworowska, Z. Skull structure in *Kennalestes* and *Asioryctes*. *Palaeontol. Pol.* **42**, 25–78 (1981).
- McKenna, M. C., Kielan-Jaworowska, Z. & Meng, J. Earliest eutherian mammal skull from the Late Cretaceous (Coniacian) of Uzbekistan. *Acta Palaeontol. Pol.* **45**, 1–54 (2000).
- Cifelli, R. L. in *Mammal Phylogeny* Vol. 1 (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 205–215 (Springer, New York, 1993).
- Rougier, G. W., Wible, J. R. & Novacek, M. J. Implications of *Deltatheridium* specimens for early marsupial history. *Nature* **396**, 459–463 (1998).
- Cifelli, R. L. & Muizon, C. de. Dentition and jaw of *Kokopellia juddi*, a primitive marsupial or near marsupial from the medial Cretaceous of Utah. *J. Mamm. Evol.* **4**, 241–258 (1997).
- Muizon, C. de. *Mayulestes ferox*, a borhyaenoid (Metatheria, Mammalia) from the early Palaeocene of Bolivia. Phylogenetic and palaeobiological implications. *Geodiversitas* **20**, 19–142 (1998).
- Luo, Z.-X., Cifelli, R. L. & Kielan-Jaworowska, Z. Dual origin of tribosphenic mammals. *Nature* **409**, 53–57 (2001).
- Luo, Z.-X., Kielan-Jaworowska, Z. & Cifelli, R. L. In quest for a phylogeny of Mesozoic mammals. *Acta Palaeontol. Pol.* **47**, 1–78 (2002).
- Krebs, B. Das Skelett von *Henkelotherium guimarotae* gen. et sp. nov. (Eupantotheria, Mammalia) aus dem Oberen Jura von Portugal. *Berliner Geowiss. Abh.* **133**, 1–110 (1991).
- Rougier, G. W. *Vincelestes neuquenianus* Bonaparte (Mammalia, Theria), un primitivo mammifero del Cretacico Inferior de la Cuenca Neuquina. Thesis, Univ. Nac. Buenos Aires (1993).
- Rich, T. H. *et al.* A tribosphenic mammal from the Mesozoic of Australia. *Science* **278**, 1438–1442 (1997).
- Rich, T. H. *et al.* A second tribosphenic mammal from the Mesozoic of Australia. *Records Queen Victoria Mus.* **110**, 1–9 (2001).
- Kielan-Jaworowska, Z. Postcranial skeleton in *Kennalestes* and *Asioryctes*. *Palaeontol. Pol.* **37**, 65–83 (1977).

31. Kielan-Jaworowska, Z. Postcranial skeleton in Zalambdalestidae. *Palaeontol. Pol.* **38**, 3–41 (1978).
32. Horowitz, I. The tarsus of *Ukhaatherium nessovi* (Eutheria Mammalia) from the Late Cretaceous of Mongolia: an appraisal of the evolution of the ankle in basal therians. *J. Vert. Paleontol.* **20**, 547–560 (2000).
33. Szalay, F. S. & Trofimov, B. A. The Mongolian Late Cretaceous *Asiatherium*, and the early phylogeny and paleobiogeography of Metatheria. *J. Vert. Paleontol.* **16**, 474–509 (1996).
34. Szalay, F. S. *Evolutionary History of the Marsupials and an Analysis of Osteological Characters* 1–481 (Cambridge Univ. Press, Cambridge, 1994).
35. Argot, C. Functional-adaptive anatomy of the forelimb in the Didelphidae, and the paleobiology of the Paleocene marsupials *Mayulestes ferox* and *Pucadelphis andinus*. *J. Morph.* **247**, 51–79 (2001).
36. Kielan-Jaworowska, Z. & Gambaryan, P. P. Postcranial anatomy and habits of Asian multituberculate mammals. *Fossils Strata* **36**, 1–92 (1994).
37. Krause, D. W. & Jenkins, F. A. Jr The postcranial skeleton of North American multituberculates. *Bull. Mus. Comp. Zool.* **150**, 199–246 (1983).
38. Rowe, T. B. Definition, diagnosis, and origin of Mammalia. *J. Vert. Paleontol.* **8**, 241–264 (1988).
39. Meng, J. & Wyss, A. R. Multituberculate and other mammal hair recovered from Palaeogene excreta. *Nature* **385**, 712–714 (1997).
40. Beard, K. C. in *Primates and Their Relatives in Phylogenetic Perspective* (ed. MacPhee, R. D. E.) 63–90 (Plenum, New York, 1992).
41. Szalay, F. Z. & Lucas, S. G. The postcranial morphology of Paleocene *Chriacus* and *Mixodectes* and the phylogenetic relationships of archontan mammals. *New Mexico Mus. Nat. Hist. Sci. Bull.* **7**, 1–47 (1996).
42. McLeod, N. & Rose, K. D. Inferring locomotory behaviour in Paleogene mammals via eigenshape analysis. *Am. J. Sci.* **293**, 300–355 (1993).
43. Jenkins, F. A. Jr in *Primate Locomotion* (ed. Jenkins, F. A. Jr) 85–116 (Academic, New York, 1974).
44. Schilling, N. & Fischer, M. S. Kinematic analysis of treadmill locomotion of tree shrews, *Tupaia glis* (Scandentia: Tupaiidae). *Z. Saugetierkd.* **64**, 129–153 (1999).
45. Loope, D. B., Dingus, L., Swisher, C. C. III & Minjin, C. Life and death in a Late Cretaceous dunefield, Nemegt Basin Mongolia. *Geology* **26**, 27–30 (1998).
46. Murphy, W. J. *et al.* Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**, 2348–2351 (2001).
47. Eizirik, E., Murphy, W. J. & O'Brien, S. J. Molecular dating and biogeography of early placental

mammal radiation. *J. Hered.* **92**, 212–219 (2001).

48. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998).
49. Hedges, S. B. & Kumar, S. Technical comments: divergence times of eutherian mammals. *Science* **285**, 2031a (1999).
50. Foote, M., Hunter, J. P., Janis, C. M. & Sepkoski, J. Jr Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. *Science* **283**, 1310–1314 (1999).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank K.-Q. Gao, M. Ellison, S.-A. Ji, M. A. Norell and D. Ren for collaborative field work; J. D. Archibald, R. L. Cifelli, Z. Kielan-Jaworowska, M. J. Novacek and G. W. Rougier for sharing ideas on early mammal research; K. C. Beard, M. Fischer, D. Gebo, M. Sánchez-Villagra and F. S. Szalay for discussions on limb anatomy and reconstructing locomotory patterns of fossil mammals; M. R. Dawson and A. Weil for improving the paper; A. Henrici for her skilful preparation; and M. Klingler for illustration of Fig. 1. We received funding from the Ministry of Land Resources of People's Republic of China and National Natural Science Foundation of China (Q.J.), National Science Foundation of USA and National Geographic Society (Z.-X.L.), and the Netting/O'Neil Funds of Carnegie Museum (Z.-X.L. and J.R.W.), and Brackenridge Fellowship of the University of Pittsburgh (J.A.G.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Z.-X.L. (e-mail: luoz@carnegiemuseums.org).

The acceleration of cosmic-ray protons in the supernova remnant RX J1713.7–3946

R. Enomoto*, T. Tanimori†, T. Naito‡, T. Yoshida§, S. Yanagita§, M. Mori*, P. G. Edwards||, A. Asahara†, G. V. Bicknell¶, S. Gunji#, S. Hara☆, T. Hara‡, S. Hayashi**, C. Itoh§, S. Kabuki*, F. Kajino**, H. Katagiri*, J. Kataoka†, A. Kawachi*, T. Kifune††, H. Kubo†, J. Kushida☆, S. Maeda**, A. Maeshiro**, Y. Matsubara‡‡, Y. Mizumoto§§, M. Moriya☆☆, H. Muraishi|||, Y. Muraki‡‡, T. Nakase¶¶, K. Nishijima¶¶, M. Ohishi*, K. Okumura*, J. R. Patterson##, K. Sakurazawa☆, R. Suzuki*, D. L. Swaby##, K. Takano☆, T. Takano#, F. Tokanai#, K. Tsuchiya*, H. Tsunoo*, K. Uruma¶¶, A. Watanabe# & T. Yoshikoshi☆☆

* Institute for Cosmic Ray Research, University of Tokyo, Kashiwa, Chiba 277-8582, Japan

† Department of Physics, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan

‡ Faculty of Management Information, Yamanashi Gakuin University, Kofu, Yamanashi 400-8575, Japan

§ Faculty of Science, Ibaraki University, Mito, Ibaraki 310-8512, Japan

|| Institute of Space and Astronautical Science, Sagami-hara, Kanagawa 229-8510, Japan

¶ Mt Stromlo and Siding Spring Observatories, Australian National University, ACT 2611, Australia

Department of Physics, Yamagata University, Yamagata, Yamagata 900-8560, Japan

☆ Department of Physics, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8551, Japan

** Department of Physics, Konan University, Kobe, Hyogo 658-8501, Japan

†† Faculty of Engineering, Shinshu University, Nagano, Nagano 380-8553, Japan

‡‡ Solar-Terrestrial Environment Laboratory, Nagoya University, Nagoya, Aichi 464-8601, Japan

§§ National Astronomical Observatory of Japan, Mitaka, Tokyo 181-8588, Japan

||| Ibaraki Prefectural University of Health Sciences, Ami, Ibaraki 300-0394, Japan

¶¶ Department of Physics, Tokai University, Hiratsuka, Kanagawa 259-1292, Japan

Department of Physics and Mathematical Physics, University of Adelaide, SA 5005, Australia

☆☆ Department of Physics, Osaka City University, Osaka, Osaka 558-8585, Japan

Protons with energies up to $\sim 10^{15}$ eV are the main component¹ of cosmic rays, but evidence for the specific locations where they could have been accelerated to these energies has been lacking². Electrons are known to be accelerated to cosmic-ray energies in supernova remnants^{3,4}, and the shock waves associated with such remnants, when they hit the surrounding interstellar medium, could also provide the energy to accelerate protons. The signature of such a process would be the decay of pions (π^0), which are generated when the protons collide with atoms and molecules in an interstellar cloud: pion decay results in γ -rays with a particular spectral-energy distribution^{5,6}. Here we report the observation of cascade showers of optical photons resulting from γ -rays at energies of $\sim 10^{12}$ eV hitting Earth's upper atmosphere, in the direction of the supernova remnant RX J1713.7–3946. The spectrum is a good match to that predicted by pion decay, and cannot be explained by other mechanisms.

RX J1713.7–3946 is a shell-type supernova remnant (SNR) that was found in the ROSAT all-sky survey⁷. Observations with the ASCA satellite have revealed intense non-thermal X-ray emission from the northwest rim⁸, and TeV γ -ray emission (1 TeV = 10^{12} eV) has been detected from the northwest rim by the CANGAROO ('collaboration of Australia and Nippon for a γ -ray observatory in the outback') 3.8-m Cerenkov telescope⁹. Infrared and radio observations have revealed the possible association of the SNR with a

molecular cloud complex¹⁰.

The CANGAROO air Cerenkov telescope, which is intended to detect very high energy γ -rays, is located near Woomera, South Australia. The 3.8-m telescope⁹, which operated from 1994 to 1998, was replaced in 2000 by a 10-m reflector with a 552-pixel camera of 0.115° square photomultiplier pixels. Observations of RX J1713.7–3946 were carried out 23–26 July and 19–27 August 2000, and 20 May–26 June 2001 with the 10-m telescope. After selecting data taken at high elevation angles ($>60^\circ$) in good weather conditions, a total of 2,332 min on-source and 1,789 min off-source data remained for further analysis.

The differential fluxes of γ -rays from RX J1713.7–3946 are plotted in Fig. 1. The number of excess events was determined from the plots of image orientation angle (α ; ref. 11) for on- and off-source runs (Fig. 1 inset) to be $3,417 \pm 240$ (14.3σ). The best fit of

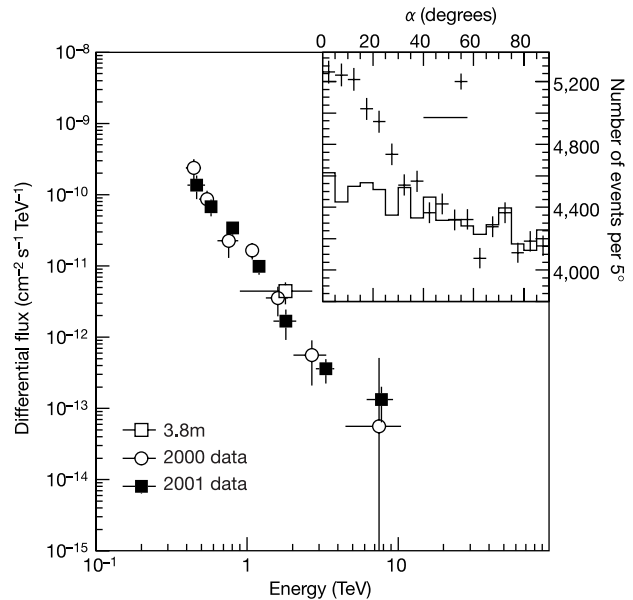


Figure 1 Differential fluxes, and distributions of the image orientation angle α (inset). Main figure: data shown by open circles and filled squares were obtained in this experiment, the open square shows data from CANGAROO-I (3.8 m)⁹. The CANGAROO-I integral flux was multiplied by 1.5 and divided by energy (TeV), assuming that the energy spectrum is proportional to $E^{-2.5}$. Inset, points with error bars show the distribution of α for on-source data, and that for off-source data is shown by the histogram. These were obtained as follows: first, 'cleaning' cuts on camera images were applied, requiring pixel pulse-heights of greater than ~ 3.3 photoelectrons, Cerenkov photon arrival times within 40 ns, and clusters of at least five adjacent triggered pixels in each event. The conventional imaging parameters¹¹ were then calculated. After applying an energy dependent distance cut, we used the shape parameters to calculate likelihoods (L ; ref. 26) under the assumptions of both γ -ray and proton origins for the cascade. The final on- and off-source data sets were obtained by excluding events with a likelihood ratio $L(\gamma - \text{ray}) / [L(\gamma - \text{ray}) + L(\text{proton})] < 0.4$ (ref. 27). The number of excess events was determined from the plots of image orientation angle α for on- and off-source runs (inset) to be $3,417 \pm 240$ (14.3σ). The γ -ray acceptance efficiency was calculated using the Monte Carlo method²⁶. In order to check our simulations, we analysed Crab nebula data obtained in December 1999 and December 2000, and found that the flux was consistent with previous experiments²⁸ to within 12%. In addition, we checked the cosmic-ray spectrum between 100 GeV and 30 TeV. The effects of night-sky background, trigger, and electronics saturation were estimated and included in uncertainties. The systematic uncertainty due to likelihood, clustering methods and pixel triggering threshold was estimated to be 20%. The energy scale uncertainty from the mirror reflectivity, mirror segment distortions, and Mie scatterings was calculated to be 15% (point to point) and 20% (overall).

the 10-m telescope results gives:

$$dF/dE = (1.63 \pm 0.15 \pm 0.32) \times 10^{-11} E^{-2.84 \pm 0.15 \pm 0.20} \quad (1)$$

where the first errors are statistical and the second are systematic with a χ^2/DOF (number degree of freedom) of 0.97, and dF/dE is in units of $\text{cm}^{-2} \text{s}^{-1} \text{TeV}^{-1}$. F and E are the gamma-ray flux and energy, respectively. When the results were fitted with a spectrum with an energy cut-off such as $\propto E^{-2} e^{-E/E_{\text{max}}}$, an E_{max} of $3.3 \pm 0.7 \text{ TeV}$ was obtained with $\chi^2/\text{DOF} = 2.3$. The fluxes are also shown in Fig. 2, multiplied by E^2 , together with multiwavelength spectra and theoretical predictions (described below). RX J1713.7–3946 is one of the intrinsically brightest galactic TeV γ -ray sources discovered, and it is notable that the power-law spectrum increases monotonically as energy decreases. This is in contrast to the spectrum of SN1006, which flattens below 1 TeV (ref. 12), consistent with synchrotron/inverse Compton (IC) models^{13–16}. Both SNRs emit X-rays via the synchrotron process, but the differing TeV spectra suggest a different emission mechanism should be acting in RX J1713.7–3946 at TeV energies.

The morphology of the TeV γ -ray emitting region is shown by the thick solid contours in Fig. 3, together with the ASCA $>2\text{-keV}$ intensity contours¹⁷ and IRAS 100- μm results. The observed TeV γ -ray intensity peak coincides with the northwest rim, and the emission extends over the ASCA contours. A possible extension towards the CO cloud in the northeast¹⁸ can also be seen.

The broadband spectrum plotted in Fig. 2 was derived using data

from ATCA (Australia Telescope Compact Array)¹⁹, ASCA^{8,17}, the EGRET²⁰ satellite-borne telescope and this work. In order to explain this spectrum, we considered three mechanisms: the synchrotron/IC process, synchrotron/bremsstrahlung, and π^0 decay produced by proton–nucleon collisions. The momentum spectra of incident particles (electrons and protons) are assumed to be:

$$dN/cdp = N_0(p/mc)^{-\alpha} \exp(-p/p_{\text{max}}) \quad (2)$$

where dN/cdp is in units of $\text{cm}^{-3} \text{TeV}^{-1}$. N is the number of particles, N_0 is the normalization factor, p is the momentum of the particle, m is the mass of the particle, c is the velocity of light, p_{max} is the maximum momentum of accelerated particles, and α is the index of the power law spectrum. The effects of acceleration limits from the age and size of the SNR are included in the exponential term. Best-fit values for the radio and X-ray fluxes due to synchrotron radiation from electrons²¹ are 2.08 for α , 126 for $p_{\text{max}} c B^{0.5}$, and 2.00 for $N_0 (V/4\pi d^2) (B)^{(\alpha+1)/2}$; here V is the volume of the radiation region, d is the distance from the Earth, $p_{\text{max}} c$ is in units of TeV, and B is in units of μG . The acceptable ranges for these parameters are 1.99–2.13, 93–139 and 4.06–0.58 respectively. The resultant best fit is plotted in Fig. 2 as the solid line.

We initially assumed the 2.7 K cosmic microwave background as the seed photons for IC scattering, where the Klein–Nishina formula was used. A similar model, applied to SN1006, matched the multiband spectrum well^{15,16}. Here, the best-fit values derived for synchrotron radiation from the incident electrons were used, and we assumed that the synchrotron and IC emission regions were the

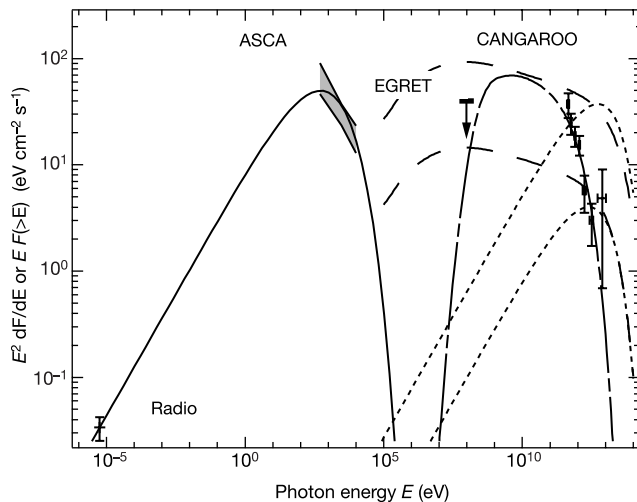


Figure 2 Multi-band emission from RX J1713.7–3946, and emission models. The radio observation was made with ATCA¹⁹. The ATCA flux at 1.36 GHz is estimated from two bright filaments lying in the northwest rim of RX J1713.7–3946 to be $S = 4 \pm 1 \text{ Jy}$ (ref. 19). The shaded area between the thick lines indicates the X-ray emission measured by the GIS detector on the ASCA satellite¹⁷. The integral flux between 0.5 and 10.0 keV was obtained from Table 4.5 in ref. 17 and the spectral index in Table 4.4. The differential flux was calculated from these two values at 3 keV, the mean GIS sensitivity. The flux uncertainty due to the extended structure of the source was considered to be within $+10\%/-30\%$, which was calculated following the procedure described in ref. 29. The EGRET upper limit corresponds to the flux of 3EG J1714–3857²⁰. The TeV γ -ray points are from this work (CANGAROO). Lines show model calculations: synchrotron emission (solid line), inverse Compton emission (dotted lines), bremsstrahlung (dashed lines) and emission from π^0 decay (short-long dashed line). Inverse Compton emission and bremsstrahlung are plotted for two cases: $3 \mu\text{G}$ (upper curves) and $10 \mu\text{G}$ (lower curves). The distance to this SNR has ambiguity as follows; the rotation velocity of the associated molecular cloud from this observation yielded a distance to the SNR of $6 \pm 1 \text{ kpc}$, in contrast to the distance of 1 kpc estimated from soft-X-ray absorption⁸. The age for the SNR is estimated to be more than 10,000 yr (for a distance of 6 kpc) or $\sim 2,000 \text{ yr}$ (for 1 kpc). Details of these models are given in the text.

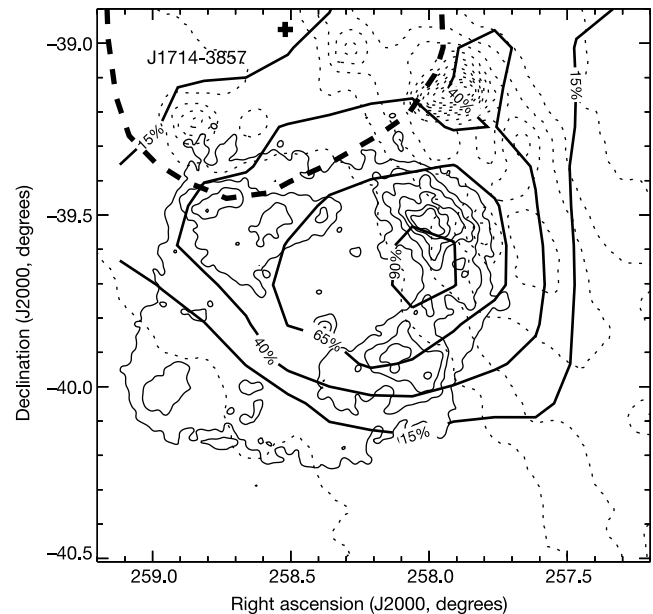


Figure 3 Profile of emission around the northwest rim of RX J1713.7–3946. The telescope tracked the point in the northwest rim of the SNR from which the maximum X-ray flux was detected (right ascension 257.9875° , declination -39.5311° in J2000 coordinates)⁸. The solid thick contours are obtained from our observations. These contours were calculated from the distribution of the detection significance determined at each location from the differences in the α plots (on- minus off-source histogram) divided by the statistical errors. The angular resolution was estimated to be 0.23° (1σ , 68% confidence level). There is a possible systematic offset of the order of 0.1° . The detection efficiency drops rapidly outside a 0.5° circle from the northwest rim, that is, the telescope pointing position. The solid thin contours are the ASCA data, and the dotted contours are IRAS 100- μm data. The sub-GeV source 3EG J1714–3857 is listed in the third EGRET catalogue²⁰ with a 95% confidence contour which includes the northeast rim (the thick dashed contour). 3EG J1714–3857 is marginally coincident with RX J1713.7–3946 and is an extended source, although there may be source confusion in this area²⁰.

same. Two representative magnetic field strengths, 3 and 10 μG , were used. The results, plotted with dotted lines in Fig. 2, show that these models are not consistent with the observed sub-TeV spectrum.

A clear feature of the synchrotron/IC process is the correlation between the peak flux and its energy. We carried out a further study of the above parameter space, taking into account the following uncertainties. For the incident-electron flux, the (1σ) fitting uncertainties were used, and we included IR emission for the IC seed photons. The IR background in the inner region of the Galaxy is not well known. We adopted a model²² that assumes a temperature of 40 K and an energy density incorporating a dilution factor. The maximum energy density was set to 1 eV cm^{-3} in order not to exceed the energy densities of cosmic rays or magnetic fields. For example, IR density at 20' apart from the nearby H II cloud (G347.61+0.20, which is the nearest and brightest IR source) was estimated to be 0.93 eV cm^{-3} using equation (3) in ref. 23. The IC calculation was carried out both with and without this IR maximum energy emission in the IC process, assuming the various incident-electron spectra. The shaded area on the right side of Fig. 4 is the resulting theoretically allowed region, with the experimentally allowed region shown by the shaded area on the left side of the figure. The predictions of synchrotron/IC models are inconsistent with our experimental data by an order of magnitude.

Although we have investigated simple cases, there is the possibility that the electron spectrum may soften at higher energies owing to synchrotron cooling. The cooling is severest in the case of the >10,000-year-old SNR for electrons exceeding E_{cool} , which is the critical energy estimated in comparison with the acceleration rate and cooling rate. If $E_{\text{cool}} < p_{\text{max}}c$, plateau-like features would

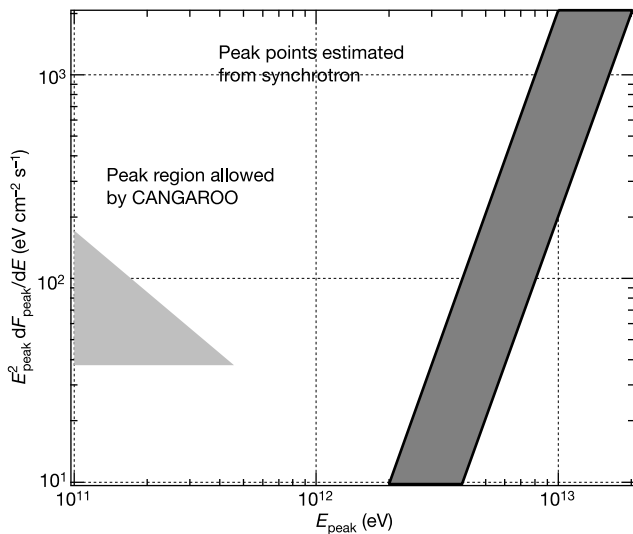


Figure 4 Allowed regions in the parameter space defined by peak flux ($E_{\text{peak}}^2 dF_{\text{peak}}/dE$) plotted against peak energy. The peak energies and fluxes for IC processes are allowed in this parameter space. The experimental flux (Fig. 2) shows the simple power-law spectrum between 400 GeV and 8 TeV, and hence the peak energy—if the IC-process is dominant—should be lower than the lowest-energy CANGAROO detection point. This corresponds to the lower-right corner of the shaded area on the left. The energy spectrum below the IC peak must flatten, and so by assuming a power-law spectral index of 2.84 for energies below 400 GeV, the experimentally allowed region is defined. On the other hand, theoretical estimates were carried out considering both the 2.7 K CMB alone, and a combination of the IR (40 K) background and the 2.7 K CMB, as target photon sources. The uncertainties from the synchrotron model fitting and those for IR (40 K) flux are taken into account. The shaded area on the right of the figure is the theoretically allowed region (1σ) under the assumption of IC process. As there is no overlap with the observationally allowed region, we can rule out IC models for the TeV γ -ray production.

appear in both the synchrotron and IC fluxes shown in Fig. 2, similar to the cases of pulsar nebulae²⁴ and blazars²⁵. These still would not fit our experimental results.

The bremsstrahlung spectrum was calculated assuming that it occurs in the same region as the synchrotron radiation. A material density of 300 protons cm^{-3} was assumed. The dashed lines in Fig. 2, for magnetic fields of 3 and 10 μG , are both inconsistent with our observation. In addition, the bremsstrahlung model is unable to simultaneously reproduce the observed GeV and TeV fluxes. Even if we neglect the EGRET data, a magnetic field of the order of 1 G with a material density of 17,500 protons cm^{-3} would be necessary. Both values are unreasonably high.

Thus electron-based models fail to explain the observational results, and so we examined π^0 decay models. The π^0 's are produced in collisions of accelerated protons with interstellar matter. A model⁶ adopting Δ -resonance and scaling was used. We adopted parameters for equation (2) of $\alpha = 2.08$ and $p_{\text{max}}c = 10$ TeV, considering the plausible parameter regions of typical shock acceleration theory. (If we simply assume the same cut-off energy for electrons, it follows from the best-fit value for the parameter $p_{\text{max}}cB^{0.5}$ of 126 that the magnetic field might be as large as 100 μG .) The result is shown by the short-long dashed curve in Fig. 2. The best-fit parameters for the total energy of accelerated protons E_0 and matter density n_0 must satisfy the equation $(E_0/10^{50})(n_0(d/6))^{-2} = 300$; here d is the distance (in kpc) to RX J1713.7–3946, E_0 is in erg, n_0 is in protons cm^{-3} . A value of $E_0 \approx 10^{50}$ erg gives n_0 of the order of 10 or 100 protons cm^{-3} for distances of 1 or 6 kpc, respectively. Both cases are consistent with the molecular column density estimated from Fig. 7 of ref. 10. Assuming a larger value of p_{max} would allow the value of $(E_0/10^{50})(n_0(d/6))^{-2}$ to be reduced. The π^0 model alone, therefore, readily explains our results, which provide observational evidence that protons are accelerated in SNRs to at least TeV energies. $\square$

Received 9 January; accepted 28 February 2002.

1. Simpson, J. A. Elemental and isotopic composition of the galactic cosmic-rays. *Annu. Rev. Nucl. Part. Sci.* **33**, 323–381 (1983).
2. Pohl, M. *Proc. 27th Int. Cosmic Ray Conf. (Hamburg) 2001, Rapporteur talk* (eds Simon, M., Lorenz, E. & Pohl, M.) (Copernicus Gesellschaft e.V., Katlenburg-Lindau, Germany, in the press); also as preprint astro-ph/0111552 at <http://xxx.lanl.gov> (2001).
3. Koyama, K. *et al.* Evidence for shock acceleration of high-energy electrons in the supernova remnant SN1006. *Nature* **378**, 255–258 (1995).
4. Tanimori, T. *et al.* Discovery of TeV gamma rays from SN 1006: Further evidence for the supernova remnant origin of cosmic rays. *Astrophys. J.* **497**, L25–L28 (1998).
5. Aharonian, F. A., Drury, L. O'C. & Völk, H. J. GeV/TeV gamma-ray emission from dense molecular clouds overtaken by supernova shells. *Astron. Astrophys.* **285**, 645–647 (1994).
6. Naito, T. & Takahara, F. High energy gamma-ray emission from supernova remnants. *J. Phys. G* **20**, 477–486 (1994).
7. Pfeffermann, E. & Aschenbach, B. in *Roentgenstrahlung from the Universe* (eds Zimmermann, H. H., Trümper, J. & Yorke, H.) 267–268 (MPE Report 263; Max-Planck-Institut fuer Extraterrestrische Physik, Garching, 1996).
8. Koyama, K. *et al.* Discovery of non-thermal X-rays from the northwest shell of the new SNR RX J1713.7–3946: The second SN 1006? *Publ. Astron. Soc. Jpn* **49**, L7–L11 (1997).
9. Muraishi, H. *et al.* Evidence for TeV gamma-ray emission from the shell type SNR RXJ1713.7–3946. *Astron. Astrophys.* **354**, L57–L61 (2000).
10. Slane, P. *et al.* Nonthermal X-ray emission from the shell-type supernova remnant G347.3–0.5. *Astrophys. J.* **525**, 357–367 (1999).
11. Hillas, A. M. in *Proc. 19th Int. Cosmic Ray Conf. (La Jolla)* Vol. 3, 445–448 (NASA Conference publication, 1985).
12. Tanimori, T. *et al.* in *Proc. 27th Int. Cosmic Ray Conf. (Hamburg)* (eds Simon, M., Lorenz, E. & Pohl, M.) Vol. 6, 2465–2468 (Copernicus Gesellschaft e.V., Katlenburg-Lindau, Germany, 2001).
13. Aharonian, F. A. & Atoyan, A. M. On the origin of TeV radiation of SN 1006. *Astron. Astrophys.* **351**, 330–340 (1999).
14. Mastichiadis, A. & de Jager, O. C. TeV emission from SN 1006. *Astron. Astrophys.* **311**, L5–L8 (1996).
15. Yoshida, T. & Yanagita, S. *Proc. 2nd INTEGRAL Workshop 'The Transparent Universe'* 85–88 (ESA SP-382 European Space Agency, Noordwijk, Netherlands, 1997).
16. Naito, T., Yoshida, T., Mori, M. & Tanimori, T. Radio to TeV gamma-ray emission from SN 1006 and shock acceleration around its rim. *Astron. Nachr.* **320**, 205–206 (1999).
17. Tomida, H. *Synchrotron Emission from the Shell-like Supernova Remnants and the Cosmic-Ray Origin* Thesis, Kyoto Univ. (1999).
18. Butt, Y. M., Torres, D. F., Combi, J. A., Dame, T. & Romero, G. Is the supernova remnant RX J1713.7–3946 a hadronic cosmic ray accelerator? *Astrophys. J.* **562**, L167–L171 (2001).
19. Ellison, D. C., Slane, P. & Gaensler, B. M. Broad-band observations and modeling of the shell-type supernova remnant G347.3–0.5. *Astrophys. J.* **563**, 191–201 (2001).

20. Hartman, R. C. *et al.* The third EGRET catalog of high-energy gamma-ray sources. *Astrophys. J. Suppl.* **123**, 79–202 (1999).
21. Rybicki, G. B. & Lightman, A. P. *Radiation Processes in Astrophysics* (Wiley and Sons, New York, 1979).
22. Mathis, J. S., Mezger, P. J. & Panagia, N. Interstellar radiation field and dust temperatures in the diffuse interstellar matter and in giant molecular clouds. *Astron. Astrophys.* **128**, 212–229 (1983).
23. Walsh, A. J., Hyland, A. R., Robinson, G. & Burton, M. G. Studies of ultracompact HII regions—I. Methanol maser survey of IRAS-selected sources. *Mon. Not. R. Astron. Soc.* **291**, 261–278 (1997).
24. Aharonian, F. A., Atoyan, A. M. & Kifune, T. Inverse Compton gamma radiation of faint synchrotron X-ray nebulae around pulsars. *Mon. Not. R. Astron. Soc.* **291**, 162–176 (1997).
25. Sikora, M., Begelman, M. C. & Rees, M. J. Comptonization of diffuse ambient radiation by a relativistic jet: The source of gamma rays from blazars. *Astrophys. J.* **421**, 153–162 (1994).
26. Enomoto, R. *et al.* Design study of CANGAROO-III, stereoscopic imaging atmospheric Cherenkov telescopes for sub-TeV gamma-ray detection. *Astropart. Phys.* **16**, 234–244 (2002).
27. Enomoto, R. *et al.* in *Proc. 27th Int. Cosmic Ray Conf. (Hamburg)* (eds Simon, M., Lorenz, E. & Pohl, M.) Vol. 6, 2477–2480 (Copernicus Gesellschaft e.V., Katlenburg-Lindau, Germany, 2001).
28. Aharonian, F. A. *et al.* The energy spectrum of TeV gamma rays from the Crab nebula as measured by the HEGRA system of imaging air Cerenkov telescopes. *Astrophys. J.* **539**, 317–324 (2000).
29. Ishizaki, Y. *Spectra and Large-scale Isotropy of the Cosmic X-ray Background from ASCA observations* Thesis (Univ. Tokyo).

Acknowledgements

We thank H. Tomida for help in understanding the ASCA observational results, M. Seta for comments on molecular cloud density, and F. Aharonian for suggestions. This work was supported by a Grant-in-Aid for Scientific Research by the Japan Ministry of Education, Science, Sports and Culture, and the Australian Research Council.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.E.
(e-mail: enomoto@icrr.u-tokyo.ac.jp).

Atomic-scale imaging of individual dopant atoms and clusters in highly *n*-type bulk Si

P. M. Voyles*, D. A. Muller*, J. L. Grazul*, P. H. Citrin*
& H.-J. L. Gossmann†

* Bell Laboratories, Lucent Technologies; and † Agere Systems, 700 Mountain Avenue, Murray Hill, New Jersey 07974, USA

As silicon-based transistors in integrated circuits grow smaller, the concentration of charge carriers generated by the introduction of impurity dopant atoms must steadily increase. Current technology, however, is rapidly approaching the limit at which introducing additional dopant atoms ceases to generate additional charge carriers because the dopants form electrically inactive clusters¹. Using annular dark-field scanning transmission electron microscopy, we report the direct, atomic-resolution observation of individual antimony (Sb) dopant atoms in crystalline Si, and identify the Sb clusters responsible for the saturation of charge carriers. The size, structure, and distribution of these clusters are determined with a Sb-atom detection efficiency of almost 100%. Although single heavy atoms on surfaces or supporting films have been visualized previously^{2–4}, our technique permits the imaging of individual dopants and clusters as they exist within actual devices.

Future generations of Si technology^{5,6} require free-electron densities in excess of 10^{21} electrons cm^{-3} . In some cases, this corresponds to concentrations of dopant atoms more than two orders of magnitude greater than their solid-solubility limit. Metastable growth techniques have significantly extended dopant concentrations beyond this limit, but the associated free-electron densities

are still found to be no higher than $\sim 5 \times 10^{20}$ electrons cm^{-3} (ref. 7). (Confining the dopants to a two-dimensional layer may help to sidestep this restriction in cases where very high-temperature processing is not required⁸.) It is therefore important to understand just how the dopant atoms become electrically inactive, leading to the observed saturation of free-electron densities.

Two generic models for electrically deactivating defect structures have been proposed for group V dopants in Si (that is, electron donors). The first model consists of between one and four dopant donor atoms surrounding a Si vacancy^{9–14}, and the second consists of only two donor atoms bound to reconfigured Si with no vacancies¹⁵. Determining which type of deactivating defect is formed in Si will establish how the fundamental limits to the free-electron concentration are reached.

We have used annular dark-field scanning transmission electron microscopy (ADF-STEM) to study highly Sb-doped Si. The samples were prepared by low-temperature molecular beam epitaxy, which provides high concentrations of Sb without introducing large amounts of Sb precipitates or Si vacancies¹⁶. The Sb density in the samples is 9.35×10^{20} atoms cm^{-3} , measured by Rutherford backscattering. The free-electron density from Hall measurements is 6.5×10^{20} electrons cm^{-3} , indicating that $\sim 30\%$ of the dopants are electrically inactive.

The STEM used here is state-of-the-art and has been optimized for very high stability during imaging¹⁷, but the real key to observing individual Sb atoms lies in the cross-section sample preparation. Simulations show that for a Si column containing a single Sb atom to be more than 25% brighter than a pure Si column, the sample must be less than 50 Å thick. Random variations in thickness must be less than the contrast of one Sb atom, and the surfaces must be free from native amorphous SiO_2 layers and damaged bulk Si layers amorphized during thinning. These rigorous requirements are achieved using chemical-mechanical polishing techniques and an etch to remove the surface oxide as described in the Methods section. We do not use ion milling due to the unavoidable damage from implantation that could be mistaken for dopant clusters.

Figure 1 is an ADF-STEM image of a cross-section of the sample in the (110) zone-axis orientation. The brightest dots on the left are atomic columns containing at least one Sb atom. The undoped region on the right shows no such very bright dots. Two images,

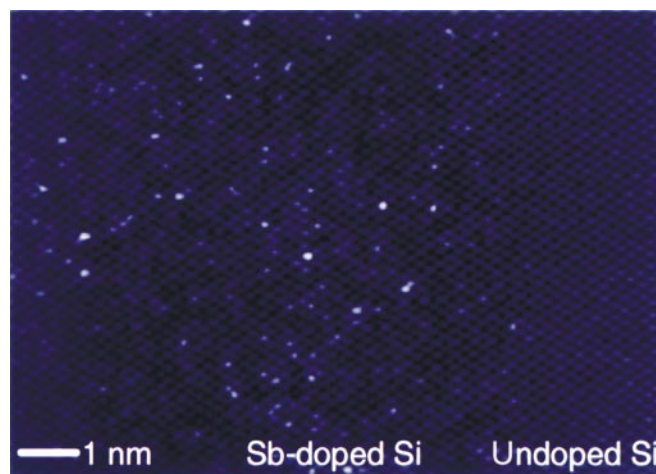


Figure 1 Annular dark-field scanning transmission electron microscopy image of a cross-section of highly Sb-doped Si. The brightest dots (left) are atomic columns containing one or more Sb atoms, absent in substrate (right), where there is no Sb. Thickness variations along the wedge have been subtracted using a second-order polynomial fit. The image has been low-pass filtered to remove scan noise, and is displayed with a nonlinear intensity scale to highlight bright features.

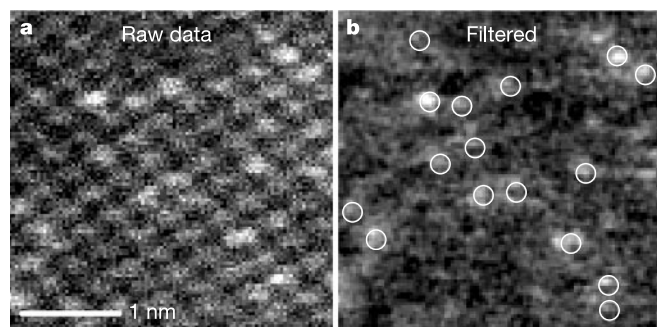


Figure 2 Segment from one of the analysed images (y) **a**, Raw data with no filtering, smoothing, or interpolation. **b**, The same region after filtering to remove the lattice. The circles indicate the Sb atoms, which are identified by the particle counting software.

described below and referred to as *x* and *y*, were obtained from Sb-doped regions of more uniform thickness. Images *x* and *y* were processed as described in the Methods section to identify the positions and intensities of the Sb-containing columns. A segment from image *y* before and after processing is shown in Fig. 2.

We find that the intensity and distribution of the Sb-occupied columns are consistent with a random substitution of Sb in Si sites, imaged with single-Sb-atom sensitivity and with almost 100% detection efficiency. For randomly distributed Sb atoms, the probability that a single column contains *s* Sb atoms is given by the binomial distribution:

$$P_{n,c}(s) = \frac{n!}{s!(n-s)!} c^s (1-c)^{n-s} \quad (1)$$

where *n* is the number of Si and Sb atoms in the column and *c* is the fractional Sb concentration in the sample. Images *x* and *y* have total areas of $127 \times 127 \text{ \AA}^2$ and $169 \times 169 \text{ \AA}^2$ and contain 1,540 and 2,747 columns, respectively. The sample thicknesses, independently measured by comparing convergent-beam electron-diffraction patterns with multi-slice electron-scattering simulations¹⁸, are 15 and 23 Å, corresponding to 9 and 11 Si atoms per column. As discussed below (and summarized in Table 1), the measured number of single-Sb-occupied columns in both images is in excellent agreement with predictions from equation (1). We also observe very few columns containing two Sb atoms, again consistent with predictions, because our samples are so thin.

ADF-STEM images are sometimes called *Z*-contrast^{19,20} because

Table 1 Number of atomic columns containing 0 to 2 Sb atoms

Number of Sb atoms	Image <i>x</i> ($120 \times 120 \text{ \AA}^2$)		Image <i>y</i> ($169 \times 169 \text{ \AA}^2$)	
	Predicted	Measured	Predicted	Measured
0	1,300	1,300 (50)	2,234	2,240 (70)
1	223	230 (30)	468	470 (40)
2	17	15 (15)	45	20 (20)

Predictions are based on equation (1). Measured values are the areas of the gaussian peaks in Fig. 3 with one standard deviation error derived from errors in the fit parameters given in parentheses. The large error in the measured values for two Sb atoms is due to the large uncertainties in the widths of the peaks.

the intensity scattered by an atom scales approximately as the atomic number $Z^{1.7}$ (close to the asymptotic Rutherford scattering limit of Z^2 ; ref. 21). Accordingly, one dopant Sb atom ($Z = 51$) should scatter about nine times more than one host Si atom ($Z = 14$). Such a large contrast might suggest that the Sb-occupied columns will exhibit discrete intensities with a distribution following equation (1), but this is not the case. The fast electron probe is strongly channelled along the atom columns^{22–24}, so the intensity scattered by a substitutional impurity depends on its depth along the column. Simulations show that both the channelling effect and the intensity of a Sb atom vary linearly with depth up to $\sim 100 \text{ \AA}$, after which they exhibit damped oscillations. Simulations also indicate that a single Sb atom is not visible in phase-contrast high-resolution TEM, even after exit-wave reconstruction (see Supplementary Information).

Figure 3 displays histograms of the column intensities from images *x* and *y*. The intensities of columns containing no Sb atoms (circles) have gaussian distributions with widths determined by variations in sample thickness and surface conditions. The Sb-containing columns (squares) also have largely gaussian distributions, but exhibit enhanced tails on the high intensity side. The green-shaded peaks have maxima at intensities (normalized to the mean unoccupied column intensity) of 1.20 and 1.16 in *x* and *y*, respectively. These values correspond to the range of intensities expected for a single Sb atom in a column. The red-shaded, high-intensity tails of the distributions can be fitted with a second gaussian of similar width, centred near twice the excess intensity of the single-Sb-atom peaks, that is, 1.40 and 1.32, respectively. The areas under the Si- and Sb-containing peaks are in extremely good agreement with the predictions of equation (1), as shown in Table 1. The detection error from noise and false positives is 3%.

Our results pass three reality checks for imaging individual atoms: (1) We count a number of dopant atoms that is consistent

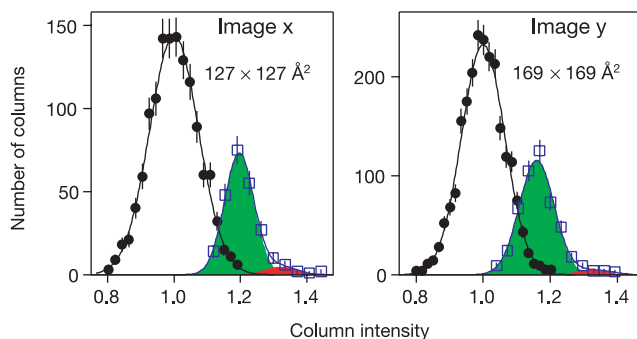


Figure 3 Histograms of atomic column intensities in two filtered images (*x* and *y*) from a highly Sb-doped Si sample. The images are different sizes and come from sample regions of different thickness. Filled circles denote columns with no Sb atoms, open squares denote columns containing Sb atoms. The error bars are $\pm \sqrt{N}$, where *N* is the number of columns in each intensity interval. Intensities are normalized by the mean unoccupied

column intensity. The green-shaded area represents columns containing one Sb atom; the red-shaded area represents columns containing two Sb atoms. The areas overlap because the Sb-column intensities depend on Sb-atom depth and because the image processing software uses a local, not global, background to detect objects.

with an independent measure of the dopant-atom concentration (Table 1). (2) We measure image intensities that agree with a statistical model for the distribution of those dopant atoms (equation (1), Fig. 3 and Table 1). (3) The features we identify as single atoms in a doped region are not seen in an undoped region imaged under identical conditions (Fig. 1). We are unaware of any

previous work on bulk-like crystals passing any of these tests.

STEM images such as those on the right-hand side of Fig. 4 contain information for identifying different deactivating dopant clusters. The clusters are denoted Sb_iV (refs 9–14) for defect structures containing a Si vacancy V surrounded by $i = 2, 3$ or 4 Sb atoms, and DP2 or DP4 for donor-pair defects formed from two Sb atoms initially occupying either second- or fourth-nearest-neighbour sites in a $\{111\}$ plane¹⁵. Our ability to identify particular defect structures is limited by two factors. First, STEM images view the structure in projection. Thus, for example, the $\langle 110 \rangle$ projections used here allow us to see only three of the four Sb atoms in a Sb_4V defect, but such an image is also consistent with one of the projections from a Sb_3V defect (see Fig. 4d). The electrical activity of a particular Sb atom cannot be firmly established for the same reason: what appears to be an inactive cluster-pair in the image could actually be two active Sb atoms at different depths. Second, the present limits of our instrumental resolution do not allow us to distinguish the small structural differences between the DP2 and Sb_2V defects (see Fig. 4a and b).

To overcome the projection problem, we have generated full three-dimensional structures using the image data and Monte Carlo simulation. Starting with the Sb-occupied column positions from the images, we populated each column with Sb atoms according to equation (1) and then placed those atoms at random depths. Nearest-neighbour Sb atoms were not allowed, consistent with X-ray absorption measurements¹⁵. The percentage of Sb atoms potentially involved in any one of the defect structures, and therefore the electrically inactive fraction, was then computed, subject to $\pm 3\%$ error from counting statistics.

We find that defects with two Sb next-nearest-neighbour atoms, which could be either Sb_2V or DP2, comprise 30 and 33% of the Sb atoms in images x and y. Both values are in excellent agreement with the Hall measurements (30%). DP4 defects account for 33 and 29%, in similarly good agreement. However, if we count both the Sb_2V or DP2 defects and the DP4 defects together, the electrically inactive fractions rise to unacceptably high values of 51 and 53%. This suggests that if DP2 and DP4 defects are both present, they cannot all be electrically inactive. Sb_3V and Sb_4V clusters are found to comprise only 1% of the Sb atoms.

From these results, we conclude that the primary deactivating defect in these samples contains only two Sb atoms. The paired defects outnumber the three- and four-atom clusters by roughly 50:1. Although our current data do not allow us to determine whether the primary pair defect is Sb_2V , DP2 or DP4, simulations indicate that atomic-resolution electron channelling experiments could tell us which defect is predominant. For example, in the DP2 structure, one of the Sb atoms is calculated to move significantly off its atomic column, while the other Sb atom remains essentially substitutional (this effect, seen in the simulation in Fig. 4a, is smaller than the scan noise in our STEM). In a Sb_2V structure, both atoms move symmetrically only a small distance towards the vacancy. Channelling draws the probe electrons tightly into the atom columns, so a Sb atom in a Sb_2V defect will scatter more strongly than the off-column Sb atom in a DP2 structure. Comparing ADF-STEM images acquired with different inner angles should highlight this effect, thereby allowing us to identify the deactivating defect. Alternatively, with the advent of aberration-corrected STEMs, these differences could be imaged directly²⁵. □

Methods

Images were collected in a JEOL 2010F-ARP STEM operated with a 200-kV electron beam, a 10-mrad probe convergence half angle, and a 50-mrad ADF inner angle¹⁷. Samples were thinned using only double-wedge mechanical polishing with diamond-coated lapping film and colloidal silica. This is similar to the tripod polishing technique²⁶, but with better control. The amorphized surface layer produced by chemical-mechanical polishing is quickly oxidized. The oxide is stripped away with a 30-s etch in 100:1 hydrofluoric acid solution immediately before we place the sample in the STEM. The result is an essentially undamaged crystal matrix for viewing.

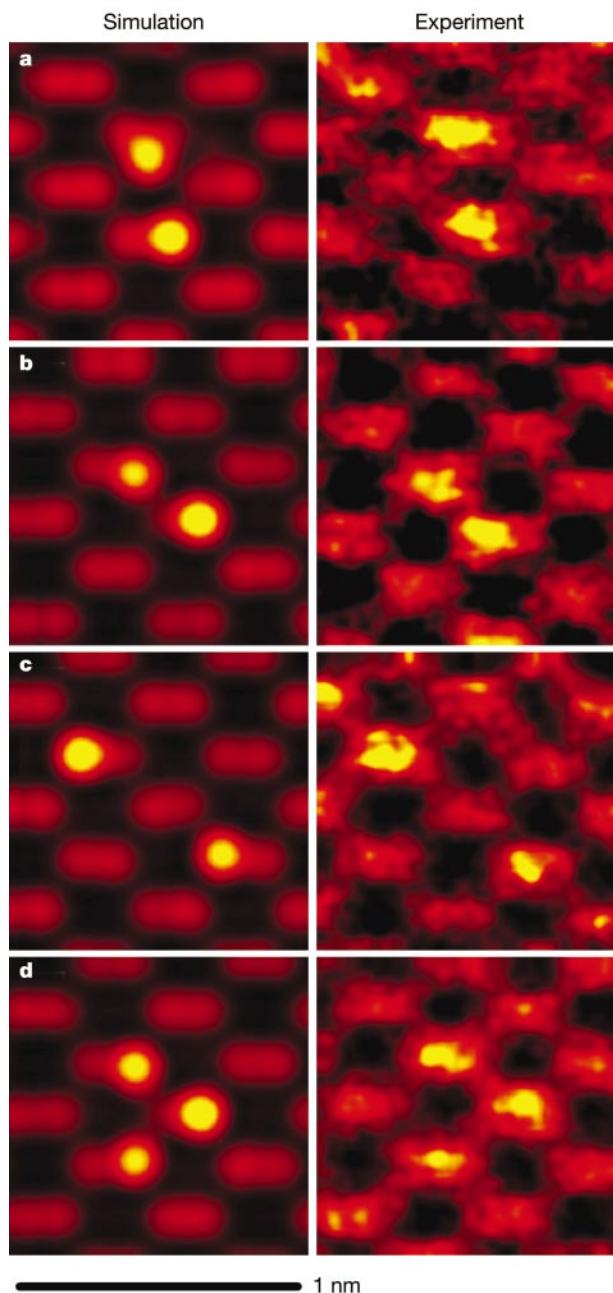


Figure 4 Simulated and experimental images of deactivating Sb clusters. The image intensity increases from black to red to yellow. Simulated images are shown on the left (see text for descriptions); **a**, DP2; **b**, DP2 viewed in a different projection; **c**, DP4; and **d**, Sb_3V . Selected experimental matches to simulated images are shown on the right. We note that (1) a Sb_2V projection also exists, similar to **a**; (2) projections of Sb_3V , and of Sb_4V with two Sb atoms in one column, each have images similar to **b**; (3) a Sb_4V projection with one doubly occupied column is similar to **d**; and (4) projections exist for DP2, Sb_2V , and DP4 with two Sb atoms in one column. Simulated images are based on structural models from ref. 15. Experimental images have been low-pass filtered to reduce noise and have had additional pixels added by interpolation (there were originally 12 pixels per column).

Image processing

The periodic contribution from the Si lattice was filtered out of the images using a singular-value decomposition. The Sb atom positions were determined from the filtered images using Source Extractor software²⁷. Source Extractor smoothed the input images with a 0.5-Å-wide gaussian, then identified collections of pixels with an area consistent with the point-spread function of the STEM (~1.6 Å wide, which is six pixels in image x and four pixels in image y) and 1.5 standard deviations (σ) per pixel above the local background as objects. The chance of a false atom detection from random noise is thus at the 9σ and 6σ levels, respectively. Consequently, the image statistics are not expected to be sensitive to precise threshold values. A 1.5σ threshold was chosen because it was the lowest value that did not produce false positives in the undoped region of Fig. 1. The detection error between 0 and 1 Sb atoms is 3%, which is largely due to remnant thickness variations.

Received 21 December 2001; accepted 28 March 2002.

1. Packan, P. A. Pushing the limits. *Science* **285**, 2079–2081 (1999).
2. Muller, E. W. Study of atomic structure of metal surfaces in the field ion microscope. *J. Appl. Phys.* **28**, 1–6 (1957).
3. Crewe, A. V., Wall, J. & Langmore, J. Visibility of single atoms. *Science* **168**, 1338–1340 (1970).
4. Nellist, P. D. & Pennycook, S. J. Direct imaging of the atomic configuration of ultradispersed catalysts. *Science* **274**, 413–415 (1996).
5. *International Technology Roadmap for Semiconductors, Update 2000* (International SEMATECH, Austin, Texas, 2000); see <http://public.itrs.net/Files/2000UpdateFinal/2kUdFinal.htm>.
6. Gossmann, H.-J., Rafferty, C. S. & Keys, P. Junctions for deep sub-100 nm MOS: How far will ion implantation take us? *Mater. Res. Soc. Symp.* **610**, B1.2.1–B1.2.10 (2000).
7. Williams, J. S. & Short, K. T. Metastable doping behavior in antimony-implanted (100) silicon. *J. Appl. Phys.* **53**, 8663–8667 (1982).
8. Citrin, P. H., Muller, D. A., Gossmann, H.-J., Vanfleter, R. & Northrup, P. A. Geometric frustration of 2D dopants in silicon: surpassing electrical saturation. *Phys. Rev. Lett.* **83**, 3234–3237 (1999).
9. Fair, R. B. & Weber, G. R. Effect of complex formation on diffusion of arsenic in silicon. *J. Appl. Phys.* **44**, 273–279 (1973).
10. Mathiot, D. & Pfister, J. C. Diffusion of arsenic in silicon: validity of the percolation model. *Appl. Phys. Lett.* **42**, 1043–1044 (1983).
11. Pandey, K. C., Erbil, A., Cargill, C. S., Boehme, R. F. & Vanderbilt, D. Annealing of heavily arsenic-doped silicon: electrical deactivation and a new defect complex. *Phys. Rev. Lett.* **61**, 1282–1285 (1988).
12. Lawther, D. W. *et al.* Vacancy generation resulting from electrical deactivation of arsenic. *Appl. Phys. Lett.* **67**, 3575–3577 (1995).
13. Ramamoorthy, M. & Pantelides, S. T. Complex dynamical phenomena in heavily arsenic doped silicon. *Phys. Rev. Lett.* **76**, 4853–4756 (1996).
14. Saarinen, K. *et al.* Identification of vacancy–impurity complexes in highly n-type Si. *Phys. Rev. Lett.* **82**, 1883–1886 (1999).
15. Chadi, D. J. *et al.* Fermi-level pinning defects in highly n-doped silicon. *Phys. Rev. Lett.* **79**, 4834–4837 (1997).
16. Gossmann, H.-J., Unterwald, F. C. & Luftman, H. S. Doping of Si thin films by low-temperature molecular beam epitaxy. *J. Appl. Phys.* **73**, 8237–8241 (1993).
17. Muller, D. A. & Grazul, J. Optimizing the environment for sub-0.2 nm scanning transmission electron microscopy. *J. Electron Microsc.* **50**, 219–226 (2001).
18. Kirkland, E. J. *Advanced Computing in Electron Microscopy* (Plenum, New York, 1998).
19. Howie, A. Image contrast and localized signal selection techniques. *J. Microsc.* **17**, 11–23 (1979).
20. Pennycook, S. J. & Narayan, J. Metastable doping behavior in antimony-implanted (100) silicon. *Appl. Phys. Lett.* **45**, 385–387 (1984).
21. Kirkland, E. J., Loane, R. F. & Silcox, J. Simulation of annular dark field STEM images using a modified multislice method. *Ultramicroscopy* **23**, 77–96 (1987).
22. Loane, R. F., Kirkland, E. J. & Silcox, J. Visibility of single heavy atoms on thin crystalline silicon in simulated annular dark field. *Acta Cryst. A* **44**, 912–927 (1988).
23. Hillyard, S. E. & Silcox, J. Thickness effects in ADF STEM zone-axis images. *Ultramicroscopy* **52**, 325–334 (1993).
24. Hillyard, S. E. & Silcox, J. Detector geometry, thermal diffuse scattering and strain effects in ADF STEM imaging. *Ultramicroscopy* **58**, 6–17 (1995).
25. Delby, N., Krivanek, O. L., Nellist, P. D., Batson, P. E. & Lupini, A. R. Progress in aberration-corrected scanning transmission electron microscopy. *J. Electron Microsc.* **50**, 177–185 (2001).
26. Klepeis, S. J., Benedict, J. P. & Anderson, R. M. in *Mater. Res. Soc. Proc.* (ed. Bravman, J. C.) 179–190 (Materials Research Society, Pittsburgh, 1988).
27. Bertin, E. & Arnouts, S. SExtractor: software for source extraction. *Astron. Astrophys. Suppl. Ser.* **117**, 393–404 (1996).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank P. Mithra for introducing us to singular value decomposition, E. Kirkland for sharing his STEM image simulation codes, and D. Wittman for help with the Source Extractor software.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D. A. M. (e-mail: davidm@bell-labs.com).

Molecular segregation observed in a concentrated alcohol–water solution

S. Dixit*, J. Crain*, W. C. K. Poon*, J. L. Finney† & A. K. Soper‡

* *Department of Physics and Astronomy, The University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK*

† *Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BE, UK*

‡ *ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, UK*

When a simple alcohol such as methanol or ethanol is mixed with water^{1,2}, the entropy of the system increases far less than expected for an ideal solution of randomly mixed molecules³. This well-known effect has been attributed to hydrophobic headgroups creating ice-like or clathrate-like structures in the surrounding water⁴, although experimental support for this hypothesis is scarce^{5–7}. In fact, an increasing amount of experimental and theoretical work suggests that the hydrophobic headgroups of alcohol molecules in aqueous solution cluster together^{2,8–10}. However, a consistent description of the details of this self-association is lacking^{11–13}. Here we use neutron diffraction with isotope substitution to probe the molecular-scale structure of a concentrated alcohol–water mixture (7:3 molar ratio). Our data indicate that most of the water molecules exist as small hydrogen-bonded strings and clusters in a ‘fluid’ of close-packed methyl groups, with water clusters bridging neighbouring methanol hydroxyl groups through hydrogen bonding. This behaviour suggests that the anomalous thermodynamics of water–alcohol systems arises from incomplete mixing at the molecular level and from retention of remnants of the three-dimensional hydrogen-bonded network structure of bulk water.

In spite of considerable research into the nature of alcohol–water mixtures at the molecular level (see, for example, ref. 2 and several experiments and computer simulations^{5,14–18}), there still seems to be little physical insight into the causes of the anomalous thermodynamic properties of alcohol–water mixtures. To investigate these matters, we perform a detailed determination of the methyl–methyl, methyl–water and water–water correlations in a concentrated mixture of methanol in water (7 methanol molecules:3 water molecules), using neutron diffraction with hydrogen isotope labelling. The diffraction data are interpreted by empirical potential structure refinement (EPSR; see Methods). Typical EPSR fits to our diffraction data on the methanol–water mixture are shown in Fig. 1. From the EPSR molecular ensembles, we derive site–site radial distribution functions (RDFs), $g_{\alpha\beta}(r)$, which describe the relative density of one type of atom, β , as a function of distance, r , from another type, α . (In this experiment, the labels α and β can each take on one of six possible atom types. They are C for methanol carbon, O for methanol oxygen, M for methyl hydrogen, H for hydroxyl hydrogen, O_W for water oxygen, and H_W for water hydrogen.) By comparing RDFs derived from our data with RDFs derived from previous diffraction data on pure methanol¹⁹ and pure water²⁰, we examine how the local ordering of the methanol molecules is modified by the addition of small amounts of water. Although many of the individual site–site RDFs are not determined separately in the diffraction experiment (for example the O_W – O_W distribution), the EPSR procedure ensures that all the estimated RDFs are consistent with all the diffraction data.

The correlations between methanol molecules, as given by the $g_{CC}(r)$, $g_{MM}(r)$ and $g_{OO}(r)$ RDFs, are shown in Fig. 2 for both pure methanol and the concentrated aqueous mixture. Two notable features emerge: when water is added, the near-neighbour distance of the CC and MM functions shift to a lower radius by about 2% for

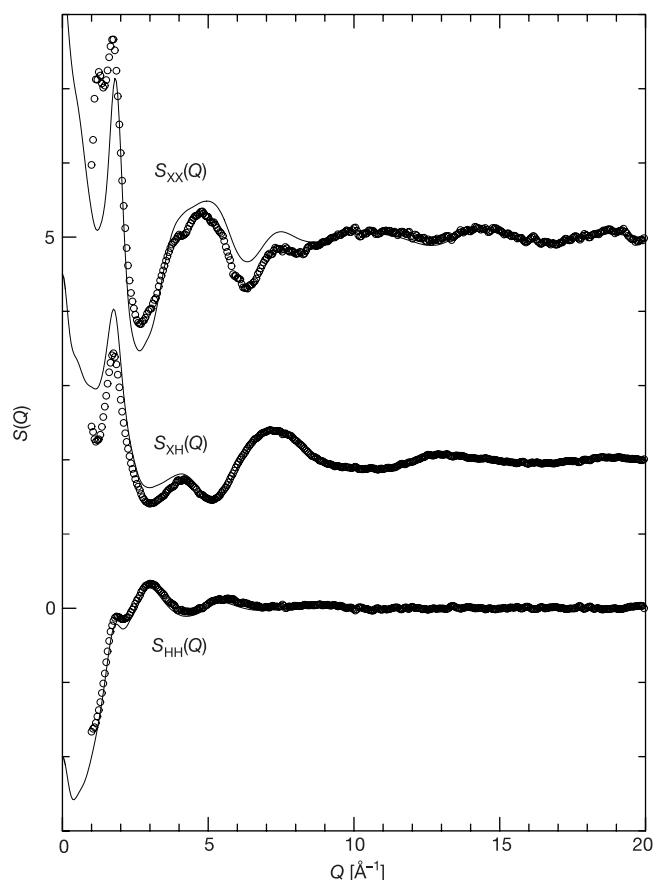


Figure 1 HH, XH and XX composite partial structure factors for a methanol–water mixture (70 mol% methanol). The case shown is where H/D substitution is performed on all hydrogen atoms (solutions (1), (6) and (7) in Methods). The measured structure factors are shown as the open circles. We note that the EPSR simulation fits simultaneously all nine composite structure factors, although only three of those fits (solid lines) are shown here. The lack of fit in some regions, particularly for the XX radial distribution function (RDF) at low wave vector transfer (Q), is believed to arise from residual nuclear recoil scattering (analogous to Compton scattering with X-rays) which is difficult to remove completely from the neutron data.

CC and more than this, by up to 20%, for MM. At the same time the first peak of the OO RDF does not move, but shrinks in size by about 50% and the second neighbour OO peak grows in amplitude and moves to a larger radius. The addition of water thus has the net effect of pressing the methyl headgroups closer together while pushing the methanol hydroxyl headgroups apart, which results in a reduction of the extent of methanol–methanol hydrogen bonding in the mixture.

The degree of hydrogen bonding in the two liquids before mixing is estimated by calculating coordination numbers for the OH and $O_W H_W$ RDFs in pure methanol and pure water respectively. (We assume here that the OH coordination number up to the first minimum in the relevant OH RDF represents the number of hydrogen bonds associated with that atom pair.) This analysis indicates that there are ~ 1.8 hydrogen bonds per methanol molecule in pure methanol and ~ 3.6 hydrogen bonds per water molecule in pure water, yielding ~ 2.3 as the average number of hydrogen bonds per molecule for a system composed of 7 parts methanol and 3 parts water that have not yet been mixed. After mixing, hydrogen bonding will be manifested in several RDFs of water and methanol: OH, OH_W , HO_W RDFs for hydrogen bonds involving methanol molecules, and $O_W H_W$, $O_W H$, and $H_W O$ RDFs

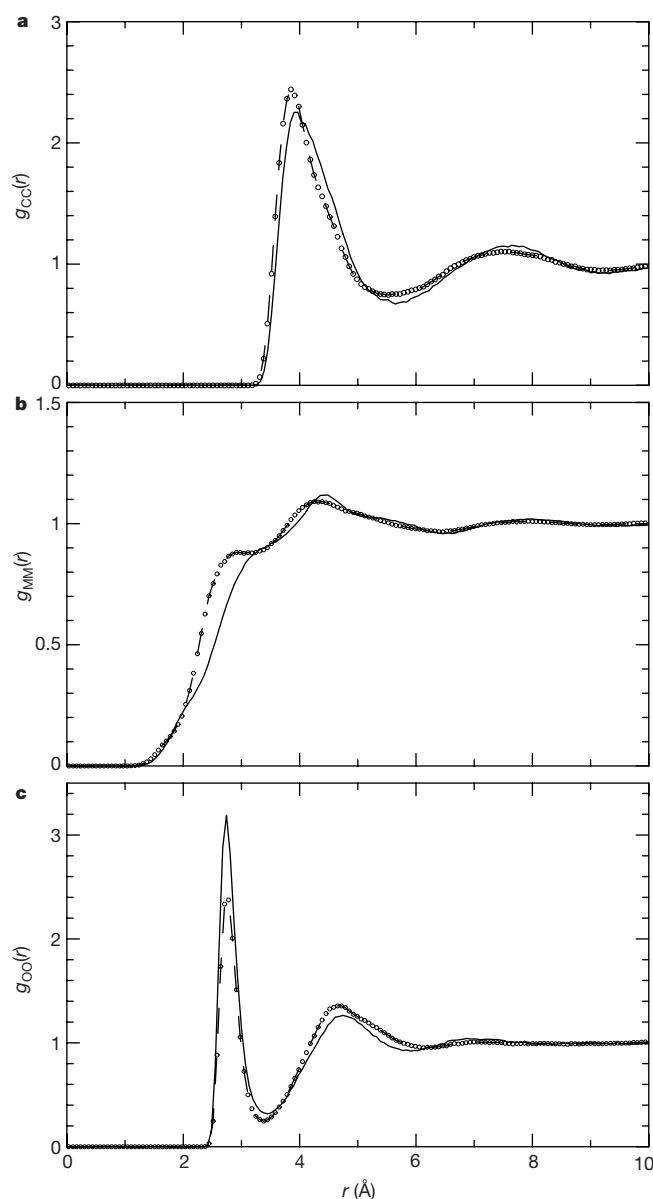


Figure 2 Methanol–methanol site–site radial distribution functions. **a–c**, The CC (**a**), MM (**b**), and OO (**c**) intermolecular methanol–methanol RDFs estimated from the structure refinement procedure for the methanol–water (7:3) mixture (dashed line, circles) are compared to the same functions for pure methanol (solid line). r , distance between atoms.

for hydrogen bonds involving water molecules. In the 7:3 methanol–water mixture, we find ~ 1.2 methanol–methanol hydrogen bonds per methanol molecule but a further ~ 0.8 hydrogen bonds to water molecules, making the total number of hydrogen bonds per methanol molecule in the mixture ~ 2.0 , that is, slightly higher than for pure methanol. For each water molecule present in the mixture, we find ~ 1.0 hydrogen bonds to other water molecules and 1.9 hydrogen bonds to methanol molecules, making the total number of hydrogen bonds per water molecule ~ 2.9 , which is only 20% less than in pure water. For the mixture as a whole, the average number of hydrogen bonds per molecules is ~ 2.3 per molecule, that is, identical to the value for the same relative amounts of pure liquids before mixing. The hydrogen bonds between methanol and water molecules created upon mixing thus compensate almost exactly for the water–water and methanol–methanol hydrogen bonds lost upon mixing the pure liquids.

The water structure is analysed using the $O_W O_W$ RDF, which is

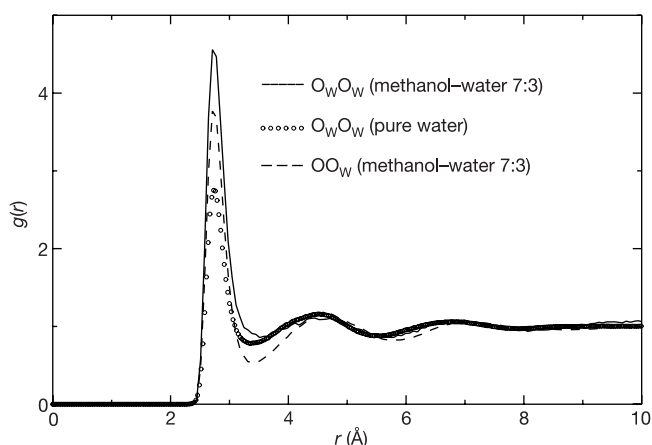


Figure 3 Water–water $O_W O_W$ radial distribution functions. The methanol–water (7:3) $O_W O_W$ RDF (solid line) is compared to the corresponding function for pure water (circles)²⁰, and with the methanol–water $O_O W$ RDF in the methanol–water (7:3) mixture (dashed line).

shown in Fig. 3 for both the mixture and pure water. The amplitude of the first peak of the $O_W O_W$ RDF for the concentrated methanol solution is larger than that of the corresponding peak for pure water, but the shapes and positions are very similar. (We note however, that the $O_W O_W$ near-neighbour coordination number, integrated out to 3.5 Å in both cases, drops from 5.0 ± 0.2 in pure water to 1.2 ± 0.2 in the methanol–water solution owing to the much lower water molecule number density in solution.) The second peak at about 4.5 Å, a signature of the hydrogen-bonded network present in pure water, is largely preserved in both position and amplitude in the concentrated alcohol mixture. Figure 3 also illustrates the correlation between water and the hydroxyl group of methanol, by showing the $O_O W$ RDF. In this function, the second-neighbour peak near 4.5 Å is sharper than in the corresponding $O_W O_W$ RDF of pure water, suggesting that the methanol hydroxyl group enhances tetrahedral structure in the water surrounding it. These results clearly contradict early predictions¹ that the water structure cannot exist in a concentrated alcohol–water solution.

A cluster analysis of the distribution of water molecules, derived from the simulation of the scattering data, provides further structural insight. If water and alcohol were randomly mixed at the molecular level, most water would exist as isolated molecules at the concentrations we have used, with distinct water clusters being rare. In practice although single water molecules (that is, cluster size 1) do occur, these constitute only about 13% of all the water molecules in the mixture. The remaining 87% water molecules occur in clusters or strings containing 2 to 20 or more molecules; the most likely structure is a three-water-molecule cluster. A snapshot of the simulation box (Fig. 4) shows groups of water molecules hydrogen-bonded together, forming small cavities in a ‘fluid’ of methyl headgroups. In this image, two water clusters can be seen left of centre and centre right, hydrogen-bonded to the hydroxyl groups of surrounding methanol molecules. The function of hydroxyl groups as bridges between methyl headgroups and water clusters explains why, on average, the methanol hydroxyl groups are further apart in the solution than in pure methanol. The concentration–temperature phase diagram of 0.7 molar aqueous methanol solution²¹ shows that the system freezes only when cooled to approximately 170 K, suggesting that the water clusters we see can be supercooled to very low temperatures.

Frank and Evans⁴ interpreted the observed excess entropies and enthalpies of solution for a wide range of solutes in terms of an ‘iceberg’ forming in the water surrounding a hydrophobic entity in aqueous solution. This appealing concept strongly influenced

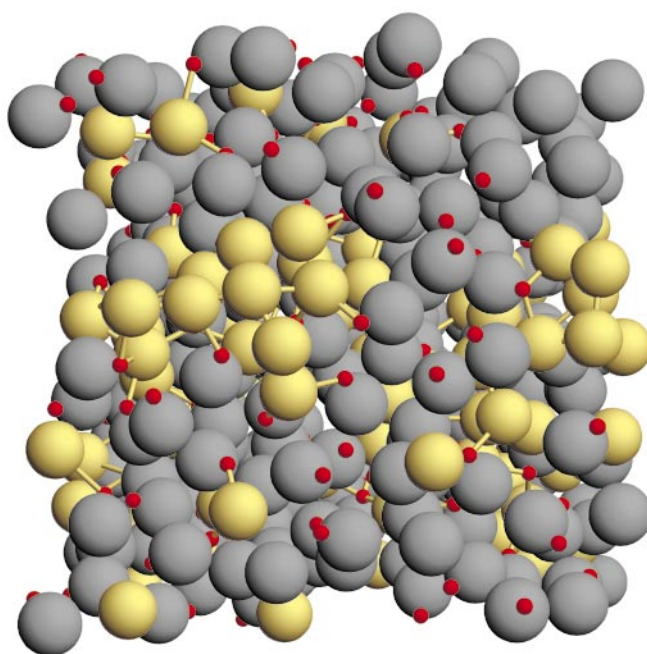


Figure 4 One of the simulated molecular boxes. Methyl groups are shown as grey spheres. Large yellow spheres have been used to highlight the positions of water molecules and small red spheres denote methanol oxygen atoms. This choice of sizes is to emphasize the fact that the average methanol carbon to water oxygen distance is almost independent of the direction in which the water molecule is viewed from the methanol molecule (see ref. 5). Yellow bonds join water oxygen atoms to other oxygen atoms within their first coordination shell ($r < 3.5$ Å).

twentieth-century thinking about the way molecules interact in aqueous solution¹¹, although it was never confirmed by structural data. The present results challenge this interpretation: contrary to speculation that water structure is either enhanced or destroyed in an alcohol solution^{1,2}, we find that the local structure of water in a concentrated methanol–water solution is surprisingly close to its counterpart in pure water. Although we cannot calculate the entropy of mixing directly from our data, they do imply that the negative excess entropy observed in these systems^{1,4} arises from incomplete mixing at the molecular level, rather than from water restructuring. We further suggest that the sign and magnitude of the excess enthalpy of mixing will be determined by an interplay between the relative strengths of the alcohol–alcohol, alcohol–water and water–water hydrogen bonds, because we find that the number of hydrogen bonds per molecule in solution is not significantly different from that in the pure liquids before mixing. The polar interaction of water with the alcohol hydroxyl group is thus likely to be a far more potent influence on the thermodynamic properties of alcohol–water mixtures than any water restructuring induced by the hydrophobic methyl groups. □

Methods

Diffraction experiment

Diffraction measurements were performed on the SANDALS diffractometer at ISIS. Seven isotopically substituted solutions (Aldrich Chemicals) were prepared: (1) CD_3OD in deuterated water, D_2O ; (2) CH_3OD in D_2O ; (3) A 50:50 mixture of solutions (1) and (2); (4) CD_3OH in H_2O ; (5) a 50:50 mixture of solutions (1) and (4); (6) CH_3OH in H_2O ; and (7) a 50:50 mixture of solutions (1) and (6). The solute–solute partial structure factor is obtained by isotope substitution on the methyl hydrogens using solutions (1), (2) and (3). Isotope substitution on the hydroxyl hydrogens of water and methanol as indicated in solutions (1), (4) and (5) gives correlations between all the hydroxyl hydrogens. Finally, solutions (1), (6) and (7) provide a measure of the correlations between all the hydrogens in the solutions, including the methanol–water correlations. Each of these three sets of solutions gives rise to three composite partial structure factors²² $S_{HH}(Q)$, $S_{XH}(Q)$ and

$S_{\text{XH}}(Q)$, where H corresponds to the labelled hydrogen and X corresponds to the remaining unlabelled atoms for each set of three solutions.

Structure refinement

The nine composite partial structure factors which resulted from analysis of the diffraction data were used as constraints in an empirical potential structure refinement (EPSR)^{23–25} simulation of the mixture, fixed at the experimental density and temperature. The cubic simulation box of side 26.77 Å contained 245 methanol and 105 water molecules. The seed potentials for the simulation were taken from the literature¹⁵. The EPSR procedure modifies these starting potential energy functions so as to make the simulated structure factors match as closely as possible the measured functions. An example of some of the fits is shown in Fig. 1. This leads to an ensemble of model molecular distributions that are consistent with the measured diffraction data.

Structure analysis

From these molecular assemblies, ensemble-averaged site–site radial distribution functions (RDFs), $g_{\alpha\beta}(r)$, can be estimated, as can other structural quantities as described below, all of which are consistent with the experimental data. Near-neighbour coordination numbers are estimated from these RDFs by integration: $N_{\alpha\beta} = 4\pi\rho_{\beta} \int_{r_{\min}}^{r_{\max}} g_{\alpha\beta}(r)r^2 dr$, where $N_{\alpha\beta}$ is the coordination number of β atoms around α atoms, ρ_{β} is the number density of β atoms in the liquid, and the range of integration is normally chosen to coincide with minima in the respective RDF. Uncertainties in these values are estimated from the observed fluctuations of $g_{\alpha\beta}(r)$ in the course of the simulation.

Cluster analysis is achieved by considering two water molecules to belong to the same cluster if they are separated by 3.5 Å or less—this is the position of the first minimum in the $O_{\text{W}}O_{\text{W}}$ radial distribution function. The size of a water cluster is then determined by counting all the water molecules that lie within the specified distance range of each other.

Received 26 October 2001; accepted 4 March 2002.

1. Franks, F. & Ives, D. J. G. The structural properties of alcohol–water mixtures. *Q. Rev.* **20**, 1–45 (1966).
2. Franks, F. & Desnoyers, J. E. in *Water Science Reviews* Vol. 1 (ed. Franks, F.) 171–232 (Cambridge Univ. Press, 1985).
3. Murrell, J. N. & Jenkins, A. D. *Properties of Liquids and Solutions* 2nd edn 102–106 (Wiley, Chichester, 1994).
4. Frank, H. S. & Evans, M. W. Free volume and entropy in condensed systems. III. Entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *J. Chem. Phys.* **13**, 507–532 (1945).
5. Soper, A. K. & Finney, J. L. Hydration of methanol in aqueous solution. *Phys. Rev. Lett.* **71**, 4346–4349 (1993).
6. Turner, J. & Soper, A. K. The effect of apolar solutes on water structure: alcohols and tetraalkylammonium ions. *J. Chem. Phys.* **101**, 6116–6125 (1994).
7. Bowron, D. T., Soper, A. K. & Finney, J. L. Temperature dependence of the structure of a 0.06 mole fraction tertiary butanol–water solution. *J. Chem. Phys.* **114**, 6203–6219 (2001).
8. Tsai, J., Gerstein, M. & Levitt, M. Keeping the shape but changing the charges: A simulation study of urea and its iso-steric analogs. *J. Chem. Phys.* **104**, 9417–9430 (1996).
9. Dixit, S., Poon, W. C. K. & Crain, J. Hydration of methanol in aqueous solutions: a Raman spectroscopic study. *J. Phys. Condens. Matter* **12**, L323–L328 (1999).
10. Wakisaka, A., Komatsu, S. & Usui, Y. Solute–solvent and solvent–solvent interactions evaluated through clusters isolated from solutions: preferential solvation in water–alcohol mixtures. *J. Mol. Liq.* **90**, 175–184 (2001).
11. D'Angelo, M., Onori, G. & Santucci, A. Self-association of monohydric alcohols in water: compressibility and infrared absorption measurements. *J. Chem. Phys.* **100**, 3107–3113 (1994).
12. Egashira, K. & Nishi, N. Low-frequency Raman spectroscopy of ethanol–water binary solution: evidence for self-association of solute and solvent molecules. *J. Phys. Chem. B* **102**, 4054–4057 (1998).
13. Yoshida, K. & Yamaguchi, T. Low temperature Raman spectroscopy of aqueous solutions of aliphatic alcohols. *Z. Naturforsch.* **A 56**, 529–536 (2001).
14. Finney, J. L., Bowron, D. T. & Soper, A. K. The structure of aqueous solutions of tertiary butanol. *J. Phys. Condens. Matter* **12**, A123–A128 (2000).
15. Ferrario, M., Haughney, M., McDonald, I. R. & Klein, M. L. Molecular dynamics simulation of aqueous mixtures: methanol, acetone and ammonia. *J. Chem. Phys.* **93**, 5156–5166 (1990).
16. Palinkas, G., Hawlicka, E. & Heinzinger, K. Molecular-dynamics simulations of water–methanol mixtures. *Chem. Phys.* **158**, 65–76 (1991).
17. Tanaka, H. & Gubbins, K. Structure and thermodynamic properties of water–methanol mixtures. *J. Chem. Phys.* **97**, 2626–2634 (1992).
18. Laaksonen, A., Kusalik, P. G. & Svishechev, I. M. Three-dimensional structure in water–methanol mixtures. *J. Phys. Chem. A* **101**, 5910–5918 (1997).
19. Yamaguchi, T., Hidaka, K. & Soper, A. K. The structure of liquid methanol revisited: a neutron diffraction experiment at -80°C and $+25^\circ\text{C}$. *Mol. Phys.* **96**, 1159–1168 (1999); Erratum **97**, 603–605 (1999).
20. Soper, A. K. The radial distribution functions of water and ice from 220 K to 673 K and at pressures up to 400 MPa. *Chem. Phys.* **258**, 121–137 (2000).
21. Murthy, S. S. N. Detailed study of ice clathrate relaxation: Evidence for the existence of clathrate structures in some water–alcohol mixtures. *J. Phys. Chem. A* **103**, 7927–7937 (1999).
22. Soper, A. K. & Luzar, A. A neutron diffraction study of dimethyl sulphoxide–water mixtures. *J. Chem. Phys.* **97**, 1320–1331 (1992).
23. Soper, A. K. Empirical potential Monte Carlo simulation of fluid structure. *Chem. Phys.* **202**, 295–306 (1996).
24. Bowron, D. T., Finney, J. L. & Soper, A. K. A structural investigation of solute–solute interactions in aqueous solutions of tertiary butanol. *J. Phys. Chem. B* **102**, 3551–3563 (1998).
25. Soper, A. K. Tests of the empirical potential structure refinement method and a new method of application to neutron diffraction data on water. *Mol. Phys.* **99**, 1503–1516 (2001).

Acknowledgements

S. D. is supported by the Edinburgh Chalmers Scholarship and the ORS award. Funding from the EPSRC Liquid Network and help from D. Bowron at an early stage of the experiments are gratefully acknowledged.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.K.S. (e-mail: a.k.soper@rl.ac.uk).

Rapid freshening of the deep North Atlantic Ocean over the past four decades

Bob Dickson*, Igor Yashayaev†, Jens Meincke‡, Bill Turrell§, Stephen Dye* & Juergen Holfort‡

* Centre for Environment, Fisheries, and Aquaculture Science, Lowestoft NR33 0HT, UK

† Bedford Institute of Oceanography, Dartmouth, Nova Scotia B2Y 4A2, Canada

‡ Institut für Meereskunde, 22529 Hamburg, Germany

§ Marine Laboratory, PO Box 101, Aberdeen AB11 9DB, UK

The overflow and descent of cold, dense water from the sills of the Denmark Strait and the Faroe–Shetland channel into the North Atlantic Ocean is the principal means of ventilating the deep oceans, and is therefore a key element of the global thermohaline circulation. Most computer simulations of the ocean system in a climate with increasing atmospheric greenhouse-gas concentrations predict a weakening thermohaline circulation in the North Atlantic as the subpolar seas become fresher and warmer^{1–3}, and it is assumed that this signal will be transferred to the deep ocean by the two overflows. From observations it has not been possible to detect whether the ocean's overturning circulation is changing, but recent evidence suggests that the transport over the sills may be slackening⁴. Here we show, through the analysis of long hydrographic records, that the system of overflow and entrainment that ventilates the deep Atlantic has steadily changed over the past four decades. We find that these changes have already led to sustained and widespread freshening of the deep ocean.

The Labrador Sea is a critical location for the Earth's climate system. In its upper and intermediate layers, annual-to-decadal variations in the production, character and thickness of its convectively formed mode water (Labrador Sea Water, LSW) directly determine the rate of the main Atlantic gyre circulation⁵. Through its deeper layers pass all of the deep and bottom waters that collectively form and drive the abyssal limb of the Atlantic meridional overturning circulation. Around its margins pass the two main freshwater flows from the Arctic Ocean to the North Atlantic (by way of the Canadian Arctic archipelago and the East Greenland shelf) that have been implicated in model experiments with a slowdown or shutdown of the meridional overturning circulation^{1–3}.

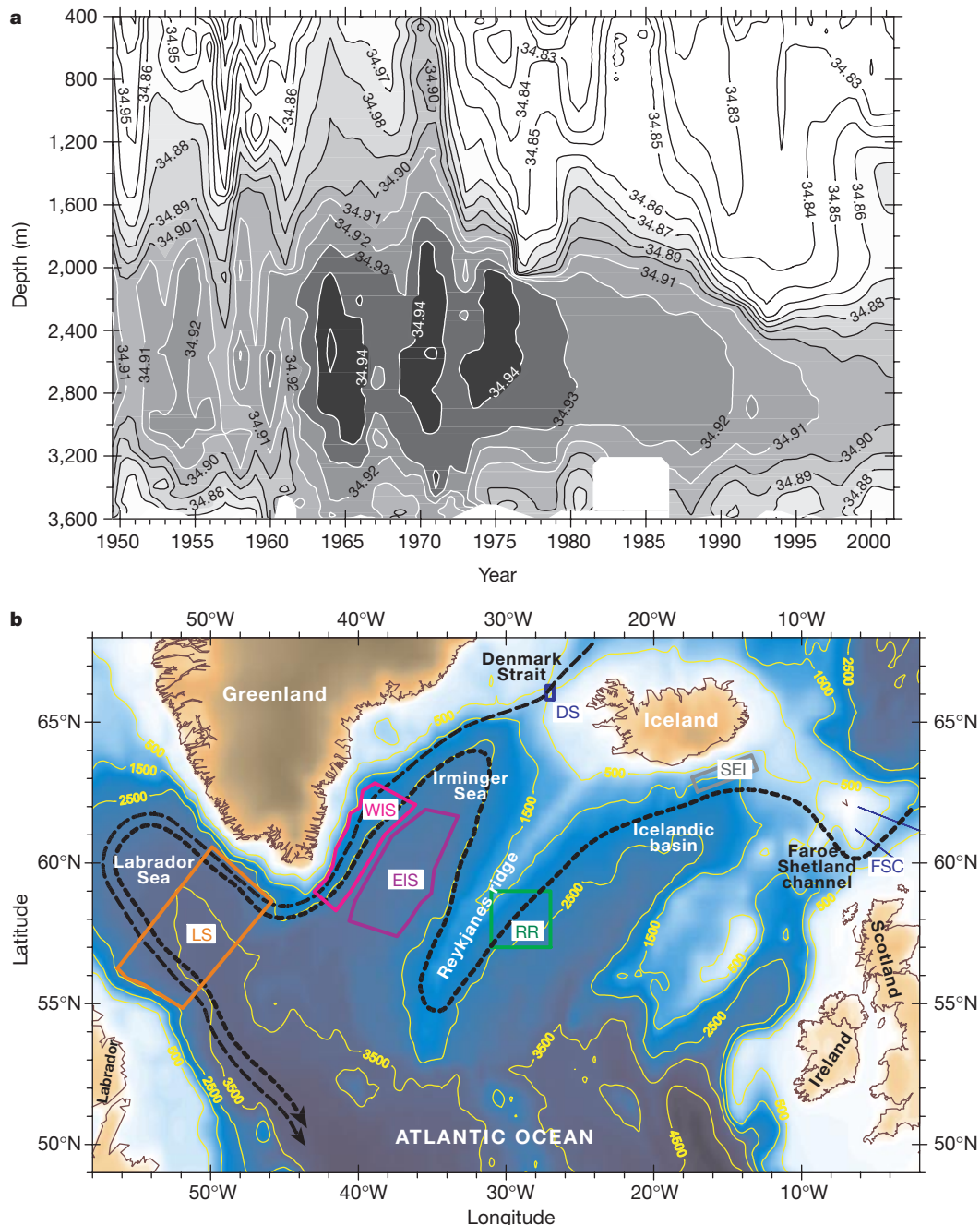
Over the past 3–4 decades, the entire water column of the Labrador Sea has undergone radical change. From 1966 to 1992, the overall cooling of the water column of the Labrador Sea was equivalent to a loss of 8 W m^{-2} continuously for 26 years. Its freshening (Fig. 1a) was equivalent to mixing-in an extra 6 m of fresh water at the sea surface⁶. As a result, the steric height (caused by

changes in the density of the water-column) in the central Labrador Sea was typically reduced by 8–10 cm. These are arguably the largest full-depth changes observed in the modern instrumental oceanographic record (see, for example, refs 7–9).

In the upper and intermediate layers of the Labrador Sea to the limit of convection (2,300 m or so), the long cooling and freshening tendency is thought to reflect the sustained if non-steady evolution of the North Atlantic Oscillation (NAO) from its most extreme negative state in the instrumental record during winters of the 1960s to its most extreme and prolonged positive state in the early 1990s¹⁰.

This evolution brought deepening convection^{11,12} and ultimately formed LSW that was fresher, colder, deeper and denser¹³ than at any other time in the history of deep measurements there.

In the upper water column (about 0–2 km), these long-sustained cooling and freshening trends were halted or partly reversed in the late 1990s (see Fig. 1a, for data post-1995, at depths between 1,000 and 2,000 m). But the deep and abyssal layers of the Labrador Sea (2,300–3,300 m) show a remarkable freshening by more than –0.01 per decade over the past 3–4 decades; the beginnings of this continuing freshening were noted 20 years ago^{14,15}. Beyond the



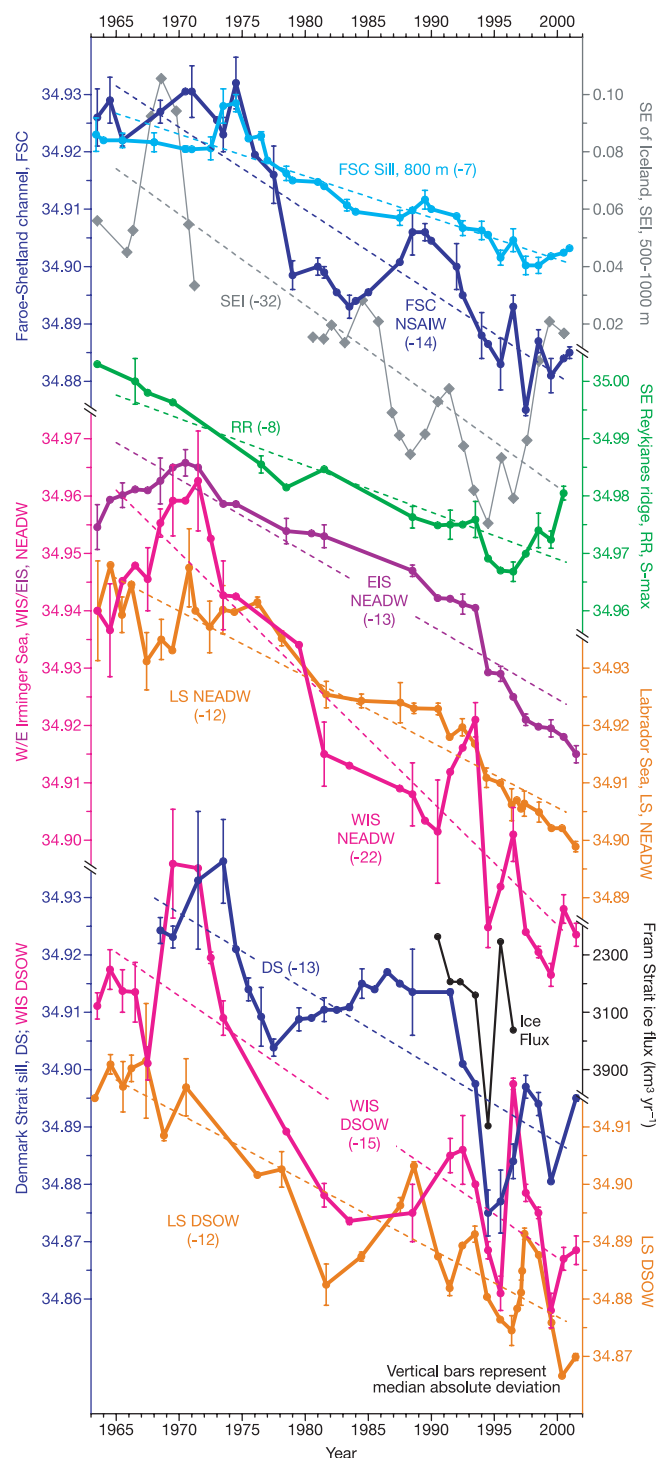


Figure 2 Evidence of sustained and rapid freshening throughout the system of overflow and entrainment that ventilates the deep Atlantic. The salinity time series shown are plotted to a common scale (with the exception of SEI at half scale), and are colour- and letter-coded to identify their locations in Fig. 1b. The mean freshening rates in p.p.m. per decade listed against each curve are calculated for the common period 1965–2000, selected as the period of the most accurate salinity data set and the period during which the North Atlantic Oscillation (NAO) exhibited a sustained change between low-index and high-index extreme states. FSC sill and FSC NSAIW describe the freshening deep outflow through the Faroe–Shetland channel at sill depth and in the Arctic Intermediate Water layer, respectively²⁰. SEI shows the depth-mean anomaly of salinity in the 500–1,000 m layer at the head of the south Icelandic basin, including data from the standard Icelandic Stokksnes section, and so represents the salinity of the water masses likely to be entrained into the eastern overflow as it leaves the Faroe Bank channel. (The short-term

spike of the 'great salinity anomaly' of the mid-1970s has been omitted). RR represents the product of that mixing, describing the salinity trend in the deep salinity maximum which marks the core of ISOW against the deep eastern flank of the Reykjanes Ridge at 57–59°N, 27–31°W. EIS NEADW, WIS NEADW and LS NEADW describe salinity time series at successive stages in the spreading of the eastern overflow as it continues into the Irminger Sea, shoals and alters along the Greenland slope and ultimately forms the Deep Water of the Labrador Sea. Curves DS, WIS DSOW and LS DSOW describe the salinity trends of the Denmark Strait overflow at the sill (depths of 500–550 m at $T < 0^{\circ}\text{C}$), in the 0–200 m near-bottom layer where the overflow plume descends the slope off southeast Greenland, and in the abyssal layers of the Labrador Sea, respectively. The observed annual mean ice flux through the Fram Strait, 1990–97, is shown for comparison¹⁷ (scale inverted) as a possible remote cause of the enhanced freshening along this path in the mid-1990s.

reach of deep convection, these changes cannot be due to local climate forcing: they probably reflect variations in the two dense overflows that renew and ventilate these deepest layers, or changes in the resident water masses that are entrained as these vigorous flows descend, or both. We consider here why the overflows freshened, and why that freshening was maintained downstream.

We propose that the ultimate source of freshening of both overflows lies in a large-scale and long-term freshening of the upper 1–1.5 km of the Nordic seas, immediately upstream. That multi-decadal change has itself been attributed to a variety of causes, many of which are also associated with the amplifying NAO: an increase in the direct export of sea ice from the Arctic Ocean^{16–18}, an increase in precipitation along the Norwegian Atlantic Current by approximately 15 cm per winter through the extension of storm activity to the Nordic seas¹⁸, and a range of factors internal to the Nordic seas¹⁹. These include an increased freshwater supply from the East Icelandic Current, a narrowing of the salty Norwegian Atlantic Current towards the coast, and a deepening of the interface between Arctic Intermediate Water and Deep Water in the Norwegian Sea^{4,19}.

Although we cannot yet partition the recent freshening of the Nordic seas into its contributory components, it is clear from some of our longest observational records that the change has occurred over a sufficiently deep layer (1–1.5 km) to affect the hydrographic character of both dense overflows crossing the Greenland–Scotland Ridge at sill depths of 700–900 m. Century-long hydrographic sections monitoring the outflow of upper Norwegian Sea Deep Water and Arctic Intermediate Water through the Faroe–Shetland channel (FSC sill and FSC NSAIW in Fig. 2) confirm that salinities have decreased there almost linearly by ~ 0.01 per decade since the mid 1970s²⁰, and our more gappy record from the deepest part of the Denmark Strait sill (DS in Fig. 2, representing salinities at 500–550 m depth for temperatures below 0 °C) indicates a very similar net freshening over the same period.

In addition to the change at both sills, Fig. 2 illustrates the salinity change at various intervals downstream where these overflows deepen, entrain, mix and spread, ultimately forming the deep and abyssal layers of the Labrador Sea (locations in Fig. 1b). All time series but one are to a common scale (the exception is the upper-ocean time series SEI, plotted at half-scale), and the bracketed values against each curve refer to the mean freshening rate over the common period 1965–2000. (Reflecting earlier usage (salinity defined in p.p.t.), a freshening rate of 0.014 per decade will be abbreviated here to ‘–14 p.p.m.’).

On leaving the Faroe Bank channel, the vigorous plume of Iceland–Scotland Overflow Water (ISOW; freshening at between –7 and –14 p.p.m. per decade; Fig. 2) will broaden, slow and deepen through entrainment of ‘resident’ water masses at the head of the south Icelandic basin before continuing south at 1,500–2,500 m depth along the eastern flank of the Reykjanes Ridge. Constructing a salinity time series for the deep salinity maximum that marks the ISOW core in this location, we find evidence of a very similar freshening rate (RR; –9 p.p.m. per decade). The same rate, period and steadiness of freshening is maintained as the ISOW-derived layer passes at depth around the Irminger Sea (there known as North East Atlantic Deep Water, thus EIS NEADW, –13 p.p.m. per decade in Fig. 2) and into the Labrador Sea (LS NEADW, –12 p.p.m. per decade).

When we similarly construct salinity series for the coldest, deepest water overflowing the Denmark Strait sill (DS, freshening at –13 p.p.m. per decade, Fig. 2), for the fully entrained DSOW-derived water mass descending the Greenland slope in the western Irminger Sea (WIS DSOW, –15 p.p.m. per decade) and for the DSOW-derived water mass occupying the deepest layers of the Labrador Sea (LS DSOW, –12 p.p.m. per decade), we find an essentially similar trend and period of freshening.

Differences in detail between the freshening trends of the two overflows are understandable if we consider that the processes contributing to the freshening of the Nordic seas are uneven in distribution. For example, Fig. 2 shows an accelerated freshening at the Denmark Strait sill (DS), in the near-bottom overflow core off southeast Greenland (WIS DSOW) and in the abyssal Labrador Sea (LS DSOW) in the mid-1990s, followed by recovery to the common trend-line, which was not experienced in the NEADW layers close by (EIS NEADW and LS NEADW). We suggest that this reflects the peak 1–2 years earlier in the efflux of ice from the Fram Strait which would more directly affect the western overflow. To illustrate this, we superimpose in Fig. 2 an observed¹⁷ ice-flux series (scale inverted).

Because both overflows are considerably altered by entrainment and mixing as they deepen and spread to the Labrador Sea²¹, we might expect their freshening signal to dilute downstream, particularly along the 5,000-km path length of the eastern overflow.

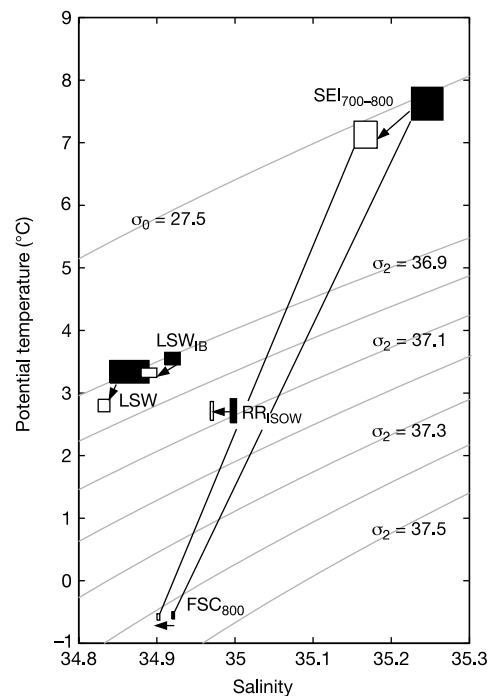


Figure 3 Maintenance of the freshening rate of Iceland–Scotland Overflow Water (ISOW) on its long spreading-path to the Labrador Sea. This freshening rate is maintained by the ISOW entraining or mixing with waters that are themselves freshening at an even greater rate. This potential temperature–salinity diagram describes the mixing relations in the South Iceland basin between overflow, entrainment and LSW, and takes account of the changing character of each of these three end-members between the 1960s (filled rectangles) and late 1990s (open rectangles). Lines of equal density, referenced adiabatically to the surface and to 2,000 dbar, are shown as σ_0 and σ_2 , respectively. Cold, dense water crossing the sill of the Faroe–Shetland channel at 800 m depth (FSC₈₀₀) will initially entrain the warm salty resident water at around that depth as it passes westward along the Iceland–Faroe slope. That end-member (SEI_{700–800}) is assigned the characteristics of water at 700–800 m depth from immediately above the permanent thermocline in this region. The product of any simple mixing between these two water-types would fall on the line connecting them, so cannot explain the relative freshness of the ISOW core (RR_{ISOW}) encountered further south against the lower flanks of the Reykjanes Ridge. That freshening can only derive from mixing with a component of LSW, the freshest end-member shown. (Note that because data to describe the changing LSW-derivative in the south Icelandic basin are sparse, we show the locus of that modified LSW (LSW_{IB}), and that of LSW at source, the latter predated by 5 years.) The present analysis suggests that RR is formed in the proportions 43% FSC₈₀₀, 31% SEI_{700–800}, and 26% LSW_{IB} although the recipe is approximate and can be expected to change with time.

So the maintenance of a more-or-less uniform freshening rate along that path (Fig. 2) is surprising. We identify two locations in particular where that freshening would be reinforced and so maintained.

The first concerns the entrainment and mixing that take place as the eastern overflow leaves the Faroe Bank channel at $>1\text{ m s}^{-1}$ to descend and decelerate through the resident water masses at the head of the south Icelandic basin²². Selecting the 500–1,000 m layer immediately southeast of Iceland as representative of the entrained water masses, Fig. 2 indicates that its mean freshening rate was more than twice as large as that of the overflow itself ($-32\text{ p.p.m. per decade}$; SEI in Fig. 2). As Fig. 3 explains, however, we need three end-members—all of them freshening with time—to account for the changing ISOW characteristics at the base of the Reykjanes Ridge. Selecting the water that overflows the Faroe–Shetland channel at sill depth (800 m) as the ‘source’ end-member, and the warm saline water mass at 700–800 m south of Iceland as the likely initial ‘entrainment’ (SEI_{700–800}; Fig. 3), we find that we cannot form the product water mass that we encounter along the lower east side of the Reykjanes Ridge (RR_{ISOW}) or explain its freshening with time unless we add the cooling and freshening influence of a sizeable component of LSW. However, the freshening rate of LSW at source was also at least as large as that of the overflow ($-13\text{ p.p.m. per decade}$, 1965–2000, for LSW defined annually as the largest volumetric temperature–salinity class). Broadly speaking, Fig. 3 implies that the overflowing source water more than doubles in volume (about $2.5\times$) through entrainment and mixing along the margins of the south Icelandic basin, and that the other two end-members (SEI_{700–800} and LSW) contribute about equally to the fully entrained product.

The divergence of curves EIS and WIS with time in Fig. 2 suggests that the NEADW layer undergoes further re-freshening in waters of the western Irminger Sea. Both curves show the salinity change on the isopycnal that lies at the core of the NEADW layer in the Irminger Sea ($\sigma_2 = 36.99\text{--}37.01$), but close to Greenland where this isopycnal shoals to 1,500–2,250 m, the freshening rate (WIS; $-22\text{ p.p.m. per decade}$) is much greater than in the centre of the basin (EIS, $-13\text{ p.p.m. per decade}$, depth range 2,120–2,550 m). We infer from this that NEADW entering from the eastern basin is re-freshened as it approaches, shoals and then passes south against the Greenland slope, either by mixing with the underlying DSOW layer, or the overlying LSW layer, or both. Thus salinities of WIS and EIS would be similar around 1970 when LSW formation was too shallow to influence that isopycnal, but diverge into the 1990s as deepening convection in the Labrador Sea carried colder and fresher LSW to a sufficient depth to influence NEADW properties in the western Irminger Sea.

We conclude then that the freshening rate of the eastern overflow was maintained downstream by mixing with waters that were themselves freshening at an equal or greater rate. A similar conclusion applies to the overflow from the Denmark Strait as it descends the Greenland slope to the abyssal Labrador Sea. As its volume rapidly doubles by entrainment south of the sill²¹, its relatively uniform freshening rate is not simply the result of a short spreading-path but must reflect mixing with overlying waters of a similar or greater rate of freshening—LSW at -13 p.p.m. or WIS NEADW at $-22\text{ p.p.m. per decade}$.

We have described a widespread, sustained, rapid and surprisingly uniform freshening of the deep and abyssal North Atlantic, south of the Greenland–Scotland Ridge, over the past four decades. Because the freshening affects both overflows and their spreading pathways downstream, it seems to confirm our supposition that subarctic change can be rapidly transferred to the deep Atlantic, and has already directly affected the abyssal limb of the Atlantic meridional overturning circulation. Other observations confirm that the deep and abyssal freshening we describe has already passed equatorward along the North American seaboard in the Deep

Western Boundary Current (W.M. Smethie, personal communication).

The question remains as to whether and to what extent these changes or these transfers reflect the onset of global change. With ‘new and stronger evidence’²³ of anthropogenic warming, coupled climate models seem to be reaching some kind of consensus that a slowdown of North Atlantic Deep Water production and of the meridional overturning circulation will be one outcome. However the issue of whether such effects are yet evident in our ocean time series remains open. Hansen *et al.*⁴ have been able to couple a moderately long, modern set of direct flow measurements to a half-century of frequent hydrography at OWS M to provide evidence of a 20% decrease in the coldest and densest part ($T < 0.3^\circ\text{C}$, $\sigma_t > 28.0$, σ_t is *in situ* density) of the overflow from the Faroe Bank channel since 1950, and such measurements as we have from the Denmark Strait give no sign of any compensating increase in the cold dense outflow through that sill²³. In the overflow hydrography that we report here, we provide the companion finding that a means exists of transferring the ‘signal’ of high-latitude climate change to the deep and abyssal headwaters of the global thermohaline circulation. However, the question of ultimate cause really hinges on whether the long amplification of the NAO, which so pervades our recent records of ocean circulation and hydrography, is itself attributable to global change; that issue is at present unresolved^{12,24}. □

Received 15 October 2001; accepted 21 March 2002.

1. Rahmstorf, S. & Ganopolski, A. Long-term global warming scenarios computed with an efficient coupled climate model. *Clim. Change* **43**, 353–367 (1999).
2. Delworth, T. L. & Dixon, K. W. Implications of the recent trend in the Arctic/North Atlantic Oscillation for the North Atlantic thermohaline circulation. *J. Clim.* **13**, 3721–3727 (2000).
3. IPCC Climate Change 2001: The Scientific Basis (eds Houghton, J. T. *et al.*) (Cambridge Univ. Press, Cambridge, 2001).
4. Hansen, B., Turrell, W. R. & Østerhus, S. Decreasing overflow from the Nordic seas into the Atlantic Ocean through the Faroe–Shetland Channel since 1950. *Nature* **411**, 927–930 (2001).
5. Curry, R. & McCartney, M. S. Ocean gyre circulation changes associated with the North Atlantic Oscillation. *J. Phys. Oceanogr.* **31**, 3374–3400 (2001).
6. Lazier, J. R. N. in *Natural Climate Variability on Decade-to-Century Time Scales* (eds Martinson, D. G. *et al.*) 295–304 (National Academy Press, Washington DC, 1995).
7. Verduin, J. & Quadfasel, D. in *European Sub-Polar Ocean Programme II, Final Scientific Report* (ed. Jansen, E.) A1, 1–11 (Univ. Bergen, Bergen, 1999).
8. Lascaratos, A., Roether, W., Nittis, K. & Klein, B. Recent changes in deep water formation and spreading in the eastern Mediterranean Sea. *Prog. Oceanogr.* **44**, 5–36 (1999).
9. Gordon, A. Weddell Deep Water variability. *J. Mar. Res.* **40**, 199–217 (1982).
10. Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science* **269**, 676–679 (1995).
11. Dickson, R. R., Lazier, J., Meincke, J., Rhines, P. & Swift, J. Long-term co-ordinated changes in the convective activity of the North Atlantic. *Prog. Oceanogr.* **38**, 241–295 (1996).
12. Hurrell, J. W. & Dickson, R. R. in *Ecological Effects of Climate Variations in the North Atlantic* (eds Stenseth, N. C., Ottersen, G., Hurrell, J. W. & Belgrano, A.) (Oxford Univ. Press, in the press).
13. Sy, A. *et al.* Surprisingly rapid spreading of newly formed intermediate waters across the North Atlantic Ocean. *Nature* **386**, 675–679 (1997).
14. Brewer, P. G. *et al.* A climatic freshening of the deep Atlantic north of 50°N over the past 20 years. *Science* **222**, 1237–1239 (1983).
15. Swift, J. H. *Climate Processes and Climate Sensitivity* (eds Hansen, J. E. & Takehashi, T.) 39–47 (AGU Geophysical Monograph 29, American Geophysical Union, Washington DC, 1984).
16. Vinje, T., Nordlund, N. & Kvambekk, A. Monitoring ice thickness in Fram Strait. *J. Geophys. Res.* **103**, 10437–10449 (1998).
17. Vinje, T. Fram Strait ice fluxes and atmospheric circulation, 1950–2000. *J. Clim.* **14**, 3508–3517 (2001).
18. Dickson, R. R. *et al.* The Arctic Ocean response to the North Atlantic Oscillation. *J. Clim.* **13**, 2671–2696 (2000).
19. Blindheim, J. *et al.* Upper layer cooling and freshening in the Norwegian Sea in relation to atmospheric forcing. *Deep-Sea Res.* **47**, 655–680 (2000).
20. Turrell, W. R., Slessor, G., Adams, R. D., Payne, R. & Gillibrand, P. A. Decadal variability in the composition of Faroe–Shetland Channel bottom water. *Deep-Sea Res.* **46**, 1–25 (1999).
21. Dickson, R. R. & Brown, J. The production of North Atlantic Deep Water: Sources, rates and pathways. *J. Geophys. Res.* **C 99**, 12319–12341 (1994).
22. Hansen, B. & Østerhus, S. North Atlantic–Nordic seas exchanges. *Prog. Oceanogr.* **45**, 109–208 (2000).
23. Giron, J. B., Sanford, T. B. & Käse, R. H. Synoptic sections of the Denmark Strait Overflow. *Geophys. Res. Lett.* **28**, 1619–1622 (2001).
24. Thompson, D. W. J., Wallace, J. M. & Hegerl, G. C. Annular modes in the extratropical circulation Part II: Trends. *J. Clim.* **13**, 1018–1036 (2000).

Acknowledgements

We thank S.-A. Malmberg and H. Valdimarsson for the use of unpublished hydro-data from the Icelandic Standard Sections Programme, and A. Clarke and J. Lazier, whose efforts provided much of the Labrador Sea data that we use here. This work was part of the data synthesis phase of the WOCE Program, of the VEINS Project of the MAST III

Programme of the European Community, and of the NOAA Consortium on the Ocean's Role in Climate of Scripps Institution of Oceanography, San Diego and Lamont-Doherty Earth Observatory, Palisades, New York; it was supported in part by the Office of Polar Programs and Geosciences Directorate of the NSF.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.R.D.
(e-mail: r.r.dickson@cefas.co.uk).

Ecosystem consequences of species richness and composition in pond food webs

Amy L. Downing* & Mathew A. Leibold

Department of Ecology and Evolution, University of Chicago, 1101 E. 57th Street, Chicago, Illinois 60637, USA

Resolving current concerns about the role of biodiversity on ecosystems calls for understanding the separate roles of changes in species numbers and of composition. Recent work shows that primary productivity often, but not always, saturates with species richness within single trophic levels^{1–8}. However, any interpretation of such patterns must consider that variation in biodiversity is necessarily associated with changes in species composition (identity)^{9–12}, and that changes in biodiversity often occur across multiple trophic levels^{13,14}. Here we present results from a mesocosm experiment in which we independently manipulated species richness and species composition across multiple trophic levels in pond food webs. In contrast to previous studies that focused on single trophic levels, we found that productivity is either idiosyncratic or increases with respect to species richness, and that richness influences trophic structure. However, the composition of species within richness levels can have equally or more marked effects on ecosystems than average effects of richness *per se*. Indirect evidence suggests that richness and associated changes in species composition affect ecosystem attributes through indirect effects and trophic interactions among species, features that are highly characteristic of natural, complex ecosystems.

The role of biodiversity in ecosystems is important both because it can reveal basic insights into the functioning of ecosystems, and because it has implications for how humans respond to current losses in global biodiversity. In a given situation, changes in biodiversity will influence local ecosystems depending on the identity^{5,15–18} and number^{4,6,8,18,19} of species going extinct. The effects of biodiversity on ecosystems may also depend on whether declines in biodiversity occur at a single trophic level compared with multiple trophic levels¹³ when ecosystem processes are influenced by the complex set of species and trophic interactions in communities^{1,20}. Here we ask how changes in species composition and richness affect ecosystems when they occur across multiple trophic levels. We focus our analysis on ecosystem attributes that indicate the importance of indirect effects of species in ecosystems.

We tested the effects of species composition and species richness in selected trophic functional groups of pond communities—macrophytes, benthic (bottom-dwelling) grazers and carnivorous

predators—while holding functional group diversity constant in field mesocosms. These functional groups were chosen because they are dominant functional groups in fishless ponds, they represent three different trophic levels, and they can be experimentally manipulated with little threat of contamination. The mesocosms also contained decomposers, phytoplankton, periphyton and zooplankton, which together with the manipulated functional groups account for the dominant functional groups in aquatic systems²¹. Although natural ponds differ from the mesocosms in some respects, the mesocosms were subject to natural fluctuations in light, temperature and rainfall, and should represent better analogues of natural systems than more controlled laboratory situations.

The experiment was designed to examine the impacts of declining richness while simultaneously estimating the relative impact of random compositional changes that are associated with biodiversity loss. To disentangle effects of richness from composition, we nested species composition treatments within diversity treatments (see Methods for details). First, we created three species richness treatments with one, three or five species per functional group. Second, within each level of richness, we nested and replicated seven unique species compositions (particular combinations of species). Testing for the effects of all possible combinations is experimentally prohibitive; however, we tested for the effects of seven random draws of species compositions from a fixed species pool within each richness level, providing an unbiased sample of possible species combinations. To draw attention to the effects of composition *per se* rather than the effects of extinction-prone species, we treat all manipulated species equally. Thus our approach does not consider how likely each of the manipulated species is to go extinct, and we emphasize that in situations of environmental concern, compositional effects will depend on additional factors that determine which species are most likely to go extinct^{13,22}. We also recognize that as a greater number of species are lost from a community, the number of possible species combinations changes and the proportional similarities between combinations may decrease.

We monitored ecosystem attributes other than the manipulated species in order to capture the importance of trophic interactions and indirect effects of the manipulated species in ecosystems. Specifically, we measured decomposition rates, and phytoplankton, periphyton and zooplankton biomass, all of which are primarily measures of the activities or abundances of unmanipulated functional groups. For example, predators consume zooplankton and benthic grazers. Zooplankton and benthic grazers in turn consume phytoplankton, periphyton, and decomposers. The ecosystem response variables also included ecosystem productivity and respiration, as calculated from diurnal oxygen cycles²³. In our system, ecosystem productivity and respiration are determined primarily by periphyton, phytoplankton and microbes (based on allometric calculations involving the metabolism and biomass of these organisms), providing a broad measure of collective metabolic activity in the larger community primarily involving trophic groups that were not manipulated.

Our results show that ecosystem productivity is greatest at the highest richness level (Fig. 1b and Table 1). Treatments with three and nine species had very similar and slower rates compared with the most diverse communities of 15 species. Unlike previous studies, we found that effects of species richness on productivity and respiration did not saturate and were synergistic³, enhanced in only the most diverse communities, at least over this range of richness levels. Patterns for respiration are similar but not quite significant (Fig. 1c and Table 1). Richness also influences the partitioning of plankton biomass within the ecosystems (Fig. 1d–f and Table 1). Phytoplankton biomass increased and zooplankton and periphyton biomass decreased with richness. In contrast, species richness did not significantly alter the total biomass of any of the manipulated functional groups (macrophytes, $F = 0.15$,

* Present address: Department of Zoology, Ohio Wesleyan University, Delaware, Ohio 43015, USA.

$P = 0.86$; grazers, $F = 1.7$, $P = 0.21$; predators, $F = 0.40$, $P = 0.673$; degrees of freedom (d.f.) = 2,18 in all cases, analysis of variance (ANOVA)). This result suggests that there was compensation among species within functional groups as diversity increased.

In contrast to richness, species composition did not significantly affect phytoplankton, periphyton and zooplankton biomass, but had significant and equally variable or more variable effects on productivity, respiration and decomposition rates, compared with richness (Fig. 2 and Table 1). Furthermore, in contrast to richness, the biomass of all the manipulated functional groups varied with composition (macrophytes, $F = 49.7$, $P < 0.001$; grazers, $F = 6.31$, $P < 0.001$; predators, $F = 6.24$, $P < 0.001$; d.f. = 18,42 in all cases, ANOVA). The effects of species composition on productivity, respiration, and decomposition rates were at least as large or larger in magnitude than the average effects of species richness (Fig. 2). Although some of these effects might be related to changes in similarity among communities, we found no relationship between pairwise compositional similarity and similarity of ecosystem functioning in this experiment (A.L.D. and J. T. Wootton, manuscript in preparation).

A central debate surrounding 'diversity–ecosystem functioning' relationships is whether effects of diversity are due to the presence of one or a few species with strong and direct effects on ecosystems (that is, the sampling effect)^{9,11,24}, or if the effects are due to diversity *per se*, involving more complex interactions between species including resource complementarity, species interactions, and indirect effects^{8,20,25}. Our data provide evidence indicating that the manipulated species are largely influencing ecosystem properties indirectly by altering abundances of other species with strong effects, and are probably not modifying ecosystem processes solely through the direct effects of a few species. First, the absence of an effect of richness on the biomass of manipulated functional groups provides no evidence for the sampling effect. Second, many of the ecosystem responses are primarily measures of unmanipulated functional groups such as phytoplankton, periphyton, zooplankton and microbes. Macrophytes, periphyton grazers and predators are known to influence ecosystem dynamics including nutrient recy-

cling dynamics, grazing rates and predation rates^{26,27}, consequently altering the biomass of unmanipulated functional or trophic groups through an abundance of direct and indirect effects. Our results suggest that not only the presence of these functional groups but also their composition and diversity may alter ecosystem functioning.

Even ecosystem productivity was probably controlled mostly by phytoplankton in our experiment, rather than the manipulated macrophyte species, which probably contributed proportionately little to oxygen dynamics. Although macrophyte biomass contributed strongly to total community biomass (42% or $2,654 \pm 315$ mg dry weight, standard error (SE)) it did not vary with species richness. In contrast, periphyton and phytoplankton biomass changes were related to species richness (Fig. 1 and Table 1) even though they contributed much less on average to community biomass (1% or 65.0 ± 4.7 mg dry weight, SE, for periphyton, and 8.7% or 553.1 ± 105.5 mg dry weight, SE, for phytoplankton). However, the effects of changes in plant biomass on ecosystem production also depend on productivity to biomass (P/B) ratios of these groups. Average macrophyte P/B ratios in the literature range between 1 and 5, whereas average phytoplankton and periphyton P/B ratios generally exceed 100 (ref. 21). Using these rough calculations, we estimate that phytoplankton contributed approximately an order of magnitude more than periphyton or macrophytes to overall productivity in these mesocosms. Consequently, the effects of species richness on productivity are much more likely to involve indirect effects of the species on phytoplankton biomass (known to vary appropriately) or phytoplankton composition, rather than on the direct effects of macrophytes on productivity.

If pond food webs are highly connected and interdependent²⁸, changes in richness or composition would alter the abundance or diversity of other, unmanipulated species or functional groups in the community through a complex set of direct and indirect effects²⁹. Theory indicates that if these species or functional groups further proceed to have strong ecosystem impacts, then originally small shifts in species composition or species richness could translate into large effects on ecosystem functioning³⁰. For example, in our experiment the effects of species richness

Table 1 Effects of species diversity and composition on ecosystem function

		Ecosystem rate						
Source of variation	d.f.	Productivity		Respiration		Decomposition		
		m.s./m.s.e.	F	m.s./m.s.e.	F	m.s./m.s.e.	F	
Diversity	2	0.0240	4.84*	0.0114	2.90§	0.0790	0.03	
	18	0.0049		0.0039		2.3100		
Time × diversity	10	0.0006	0.36	0.0002	0.27	0.4277	0.74	
	90	0.0015		0.0006		0.5770		
Composition	18	0.0049	2.59†	0.0039	5.10‡	2.3102	2.02*	
	42	0.0019		0.0008	1.1460			
Time x composition	90	0.0059	4.83*	0.0006	1.40*	0.5772	1.07	
	210	0.0012		0.0004		0.5390		
		Trophic functional group						
Source of variation	d.f.	Periphyton biomass¶		Phytoplankton biomass¶		d.f.	Zooplankton biomass¶	
		m.s./m.s.e.	F	m.s./m.s.e.	F		m.s./m.s.e.	F
Diversity	2	1.478	6.06‡	1.914	8.12†	2	1.163	5.42*
	18	0.244		0.236		18	0.214	
Time × diversity	10	0.069	1.32	0.070	0.65	4	0.250	1.17
	90	0.052		0.108		36	0.213	
Composition	18	0.244	0.59	0.236	0.99	18	0.214	1.17
	42	0.412		0.238		42	0.184	
Time × composition	90	0.052	0.81	0.108	0.66	36	0.213	1.54
	210	0.064		0.164		84	0.138	

Univariate repeated measures ANOVA results are reported for among the group effects diversity and composition. Significance values for within-group effects time × diversity and time × composition are adjusted using the Greenhouse–Geisser correction for degrees of freedom (d.f.). m.s./m.s.e.—the mean square (m.s.) and mean square error (m.s.e.) are shown as the first and second entry respectively for each source of variation. A significant time by treatment interaction corresponds to different temporal patterns between treatments.

* $P < 0.05$; † $P < 0.01$; ‡ $P < 0.001$; § $P = 0.08$; || $P = 0.06$.

¶ Data were log-transformed to better meet the assumption of normality.

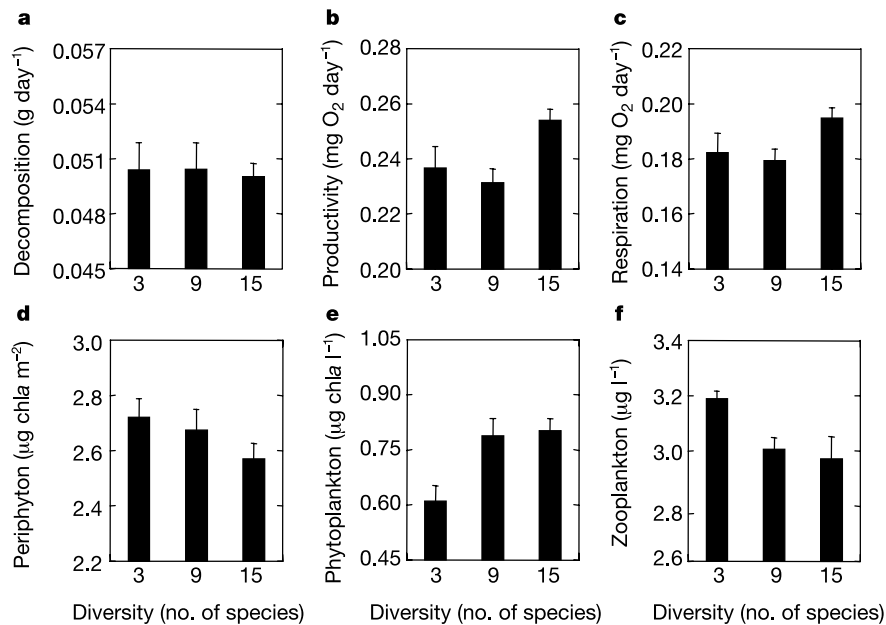


Figure 1 The response of ecosystem rates and trophic structure to species diversity. **a–c**, Ecosystem rates of decomposition (**a**), productivity (**b**) and respiration (**c**). **d–f**, Response of indicated functional group (log values): periphyton (**d**), phytoplankton (**e**) and zooplankton (**f**). Mean values for each diversity level are calculated from the seven

means of each composition level nested within diversity, averaged with respect to replicates and time. These values correspond to the main effects of the repeated measures ANOVA found in Table 1. Error bars depict standard error among the seven composition means nested within each level of diversity. Chla, chlorophyll *a*.

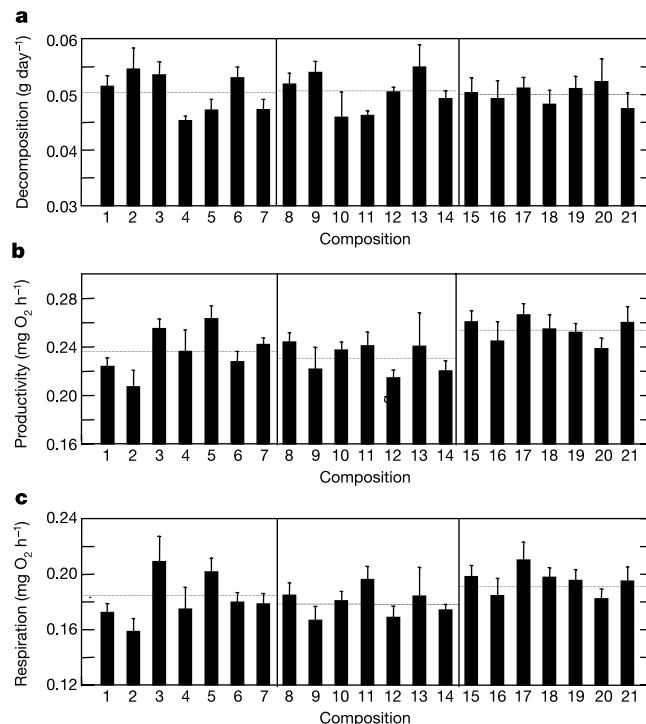


Figure 2 The response of ecosystem rates to species composition nested in species diversity. Ecosystem rates of decomposition (**a**), productivity (**b**) and respiration (**c**) are shown. Bars represent the means and standard error of the four replicates of each composition treatment averaged over time. The horizontal dotted lines indicate the overall mean for each of the three diversity levels (left, middle and right sections, respectively), corresponding to the means shown in Fig. 1a–c. All 21 unique species combinations are shown. To compare the variability in ecosystem response due to composition versus diversity, we calculated bootstrapped estimates of variance in the means of composition treatments by randomly sampling three composition treatments (equivalent to three diversity treatments) from the seven treatments nested within each level of diversity. We

calculated the mean of 150 bootstrapped estimates of variance due to composition (50 estimates within each level of diversity) to arrive at a variance estimate for each variable. Composition had equal or more variable effects than diversity on ecosystem responses, however only decomposition was significantly greater as determined by an $F_{\max}$ test (decomposition: diversity variance = 0.00000039, composition variance = 0.00000917, $F = 238$, d.f. = 2, $P < 0.01$; respiration: diversity variance = 0.000048, composition variance = 0.000135, $F = 2.8$, d.f. = 2, $P > 0.05$; productivity: diversity variance = 0.000114, composition variance = 0.000211, $F = 1.85$, d.f. = 2, $P > 0.05$).

and composition on productivity may have involved indirect mechanisms mediated through the phytoplankton assemblage. Importantly, as the potential number and types of direct and indirect effects increases rapidly with species richness, the balance of interactions may be more easily disrupted in complex food webs relative to single trophic level systems, potentially explaining our observation that species richness can have synergistic effects on productivity. (We recognize that the synergisms we see at these levels of richness may show an asymptotic relationship at yet higher levels, or may also be consistent with the 'idiosyncratic' hypothesis.) Intriguingly, results from the first major study that manipulated diversity across entire food webs in terrestrial systems also suggest an accelerating, 'synergistic' relationship between diversity and respiration¹⁹.

Our results demonstrate that both composition and species richness can alter ecosystem attributes in food webs with many trophic levels. Species richness effects appear to be synergistic or idiosyncratic, perhaps because of the greater likelihood of indirect effects than in previous studies involving single trophic levels. Additionally, our results indicate that not all ecosystem attributes respond similarly to richness and composition. Compositional effects can be as large or larger than the effects of species richness *per se*. The application of these results to environmental concerns about losses of biodiversity is incomplete because we describe ecosystem responses to random compositional change. Nevertheless, we suggest that the consequences of biodiversity loss might be complex and difficult to predict without also accounting for compositional changes that will largely depend on the factors generating non-random extinctions in nature²². □

Methods

Mesocosms

Polyethylene tanks (300 l) were filled with sand substrate and nutrient-poor well water enriched to average nitrogen and phosphorus concentrations found in local, natural ponds. We inoculated mesocosms with diverse mixtures of zooplankton, phytoplankton and periphyton from local ponds. Fibreglass screen lids prevented unwanted colonization and escape of aquatic species.

Experimental design

The species diversity treatment consisted of three, nine and 15 species evenly distributed across three functional groups: macrophytes, periphyton grazers and invertebrate predators. Seven unique species compositions were nested in each of the three levels of diversity through random draws of one, three or five species per functional group, for a total of 21 unique species composition treatments. Each composition was replicated four times (see Supplementary Information for specific species combinations for each treatment). Half of the experiment was subjected to a pulse acidification event part way through the experiment to explore the effects of pH changes on ecosystem stability, however the results of this treatment will be discussed elsewhere. Species were added in a replacement design, keeping total number of individuals (wet weight for macrophytes) in each functional group constant across diversity levels; 60 g total wet weight macrophyte biomass, 90 individual benthic grazers, and 24 individual carnivorous predators. The number of individuals per each functional group was chosen to approach species densities found in natural ponds while ensuring enough individuals of each species for reproduction. All species manipulated were observed to reproduce successfully in the mesocosms, with the exception of tadpoles and macrophytes, which exhibited growth responses. In experiments we have conducted testing for effects of starting densities on final species biomass, differences in biomass generally disappear within 8 weeks of the experiment due to reproduction, mortality, or growth of the pond biota (see Supplementary Information). Treatments were established over a six-week period, beginning with macrophyte species, followed by grazers, and lastly predators.

Ecosystem response variables

Ecosystem variables were measured six times over the course of the experiment, except zooplankton biomass, which was measured three times. We took measurements approximately every two weeks (four weeks for zooplankton) for 12 weeks, beginning four weeks after treatments were established. Ecosystem productivity and respiration rates were calculated as the difference between the maximum and minimum oxygen concentrations over a 24-h period taken with an oxygen probe. Productivity rates are the net gain of oxygen between dawn and dusk, and respiration rates are the net loss between dusk and dawn²³. We calculated decomposition rates as the loss of dry mass per day of sugar maple leaves enclosed in mosquito netting. Phytoplankton and periphyton biomass were determined from chlorophyll *a* extraction, and converted into dry weight. Zooplankton were sieved through an 80-µm mesh filter and transferred to a pre-weighed glass fibre filter for dry weight determination. To correct for phytoplankton biomass, a subsample of each

zooplankton sample was analysed for chlorophyll, which was then converted into dry weight of phytoplankton and subtracted from the zooplankton sample. All manipulated species were counted and weighed at the end of the experiment, and converted to dry weight based on length-weight regressions or direct weighing. Mean biomass and mean number of individuals of each species for each composition treatment at the termination of the experiment can be found in the Supplementary Information. Community biomass was calculated as the total biomass of the manipulated species in addition to phytoplankton, periphyton and zooplankton biomass. The average biomass of periphyton, phytoplankton and zooplankton biomass per tank was calculated by using the appropriate conversion factors (total surface area available for periphyton growth, and total volume of water per mesocosm for zooplankton and phytoplankton) to convert biomass per area or volume to biomass per mesocosm.

Statistical analyses

Statistical analyses use the appropriate error terms for the full experimental design, although the results of the acidification treatment are not reported here. Data were analysed using mixed model, repeated measures ANOVA, where species composition is a random effect nested in species diversity, which is a fixed effect. Levene's test confirmed that variances were homogeneous across diversity treatments for all response variables except zooplankton biomass ($P = 0.03$), where variance increased as the mean decreased and as compositional similarity presumably increased within the high richness treatment.

Received 27 September 2001; accepted 17 January 2002.

- Chapin, F. S. III *et al.* Consequences of changing biodiversity. *Nature* **405**, 234–242 (2000).
- Schläpfer, E., Schmid, B. & Seidl, I. Expert estimates about effects of biodiversity on ecosystem processes and services. *Oikos* **84**, 346–352 (1999).
- Schwartz, M. W. *et al.* Linking biodiversity to ecosystem function: implications for conservation ecology. *Oecologia* **122**, 297–305 (2000).
- Hector, A. *et al.* Plant diversity and productivity experiments in European grasslands. *Science* **286**, 1123–1127 (1999).
- Symstad, A. J., Tilman, D., Willson, J. & Knops, J. M. H. Species loss and ecosystem functioning: effects of species identity and community composition. *Oikos* **81**, 389–397 (1998).
- Tilman, D., Wedin, D. & Knops, J. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **379**, 718–720 (1996).
- Tilman, D. *et al.* The influence of functional diversity and composition on ecosystem processes. *Science* **277**, 1300–1302 (1997).
- van der Heijden, M. G. A. *et al.* Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**, 69–72 (1998).
- Huston, M. A. Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia* **110**, 449–460 (1997).
- Grime, J. P. Biodiversity and ecosystem function: the debate deepens. *Science* **277**, 1260–1261 (1997).
- Aarssen, L. W. High productivity in grassland ecosystems: effected by species diversity or productive species? *Oikos* **80**, 183–184 (1997).
- Huston, M. A. *et al.* No consistent effect of plant diversity on productivity. *Science* **289**, 1255a (2000).
- Petchev, O. L., McPhearson, P. T., Casey, T. M. & Morin, P. J. Environmental warming alters food-web structure and ecosystem function. *Nature* **402**, 69–72 (1999).
- Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. Quantifying threats to imperiled species in the United States. *BioScience* **48**, 607–615 (1998).
- Hooper, D. & Vitousek, P. M. The effects of plant composition and diversity on ecosystem processes. *Science* **277**, 1302–1305 (1997).
- Mikola, J. Effects of microbivore species composition and basal resource enrichment on trophic-level biomasses in an experimental microbial-based soil food web. *Oecologia* **117**, 396–403 (1998).
- Norberg, J. Resource-niche complementarity and autotrophic compensation determines ecosystem-level responses to increased cladoceran species richness. *Oecologia* **122**, 264–272 (2000).
- McGrady-Steed, J., Harris, P. M. & Morin, P. J. Biodiversity regulates ecosystem predictability. *Nature* **390**, 162–165 (1997).
- Naem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).
- Naem, S., Hahn, D. & Schuurman, G. Producer-decomposer co-dependency influences biodiversity effects. *Nature* **403**, 762–764 (2000).
- Wetzel, R. G. *Limnology* (Saunders College Publishing, Orlando, 1983).
- Pimm, S. L., Russell, G. J., Gittleman, J. L. & Brooks, T. M. The future of biodiversity. *Science* **269**, 347–350 (1995).
- Lund, O. T. *Handbook of Common Methods in Limnology* (Mosby Company, St Louis, 1979).
- Tilman, D., Lehman, C. L. & Thomson, K. T. Plant diversity and ecosystem productivity: theoretical considerations. *Proc. Natl Acad. Sci. USA* **94**, 1857–1861 (1997).
- de Ruiter, P. C., Neutel, A.-M. & Moore, J. C. Energetics, patterns of interaction strengths, and stability in real ecosystems. *Science* **269**, 1257–1260 (1995).
- Sand-Jensen, K. & Borum, J. Interactions among phytoplankton, periphyton, and macrophytes in temperate freshwaters and estuaries. *Aquat. Bot.* **41**, 137–175 (1991).
- McCollum, E. W., Crowder, L. B. & McCollum, S. A. Complex interactions of fish, snails, and littoral zone periphyton. *Ecology* **79**, 1980–1994 (1998).
- Martinez, N. D. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monogr.* **61**, 367–392 (1991).
- Wootton, J. T. The nature and consequences of indirect effects in ecological communities. *Annu. Rev. Ecol. Syst.* **25**, 443–466 (1994).
- Yodanis, C. L. *et al.* *Integration of Patterns and Dynamics* (eds Polis, G. A. & Winemiller, K. O.) 192–200 (Chapman and Hall, New York, 1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank S. Naem, J. T. Wootton, J. Bergelson, C. Pfister and M. Huston for comments on the manuscript, and J. Chase and J. Shurin for help in the field. The paper is a Kellogg Biological Station contribution, and was supported by grants from the National Science Foundation, US Department of Education, and the University of Chicago.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials and should be addressed to A.L.D. (e-mail: aldowin@owu.edu).

Low host specificity of herbivorous insects in a tropical forest

Vojtech Novotny*, Yves Basset†, Scott E. Miller‡, George D. Weiblen§, Birgitta Bremer||, Lukas Cizek* & Pavel Drozd¶

* Institute of Entomology, Czech Academy of Sciences and Biological Faculty, University of South Bohemia, Branisovska 31, 370 05 Ceske Budejovice, Czech Republic

† Smithsonian Tropical Research Institute, Apartado 2072, Balboa, Ancon, Panama

‡ Department of Systematic Biology, National Museum of Natural History, Smithsonian Institution, Washington DC 20560-0105, USA

§ Department of Plant Biology, University of Minnesota, 220 Biological Sciences Center, 1445 Gortner Avenue, St Paul, Minnesota 55108-1095, USA

|| Bergius Foundation at the Royal Swedish Academy of Sciences, PO Box 50017, SE-104 05 Stockholm, Sweden

¶ University of Ostrava, Department of Biology, 30. dubna 22, 701 03 Ostrava, Czech Republic

Two decades of research^{1–4} have not established whether tropical insect herbivores are dominated by specialists or generalists. This impedes our understanding of species coexistence in diverse rainforest communities. Host specificity and species richness of tropical insects are also key parameters in mapping global patterns of biodiversity^{1,4,5}. Here we analyse data for over 900 herbivorous species feeding on 51 plant species in New Guinea and show that most herbivorous species feed on several closely related plant species. Because species-rich genera are dominant in tropical floras, monophagous herbivores are probably rare in tropical forests. Furthermore, even between phylogenetically distant hosts, herbivore communities typically shared a third of their species. These results do not support the classical view that the coexistence of herbivorous species in the tropics is a consequence of finely divided plant resources; non-equilibrium models of tropical diversity⁶ should instead be considered. Low host specificity of tropical herbivores reduces global estimates of arthropod diversity from 31 million (ref. 1) to 4–6 million species. This finding agrees with estimates based on taxonomic collections, reconciling an order of magnitude discrepancy between extrapolations of global diversity based on ecological samples of tropical communities with those based on sampling regional faunas^{7,8}.

Host specificity is difficult to measure, and the limitations of existing studies include sampling only certain taxonomic groups rather than entire guilds, or sampling limited numbers of host plant species and lineages. Studies are often of insufficient duration, producing samples too small for quantitative analysis, or insects are sampled destructively, which precludes feeding experiments and the study of immature stages. Further, previous studies² failed to consider the phylogenetic relationships of host plants by using

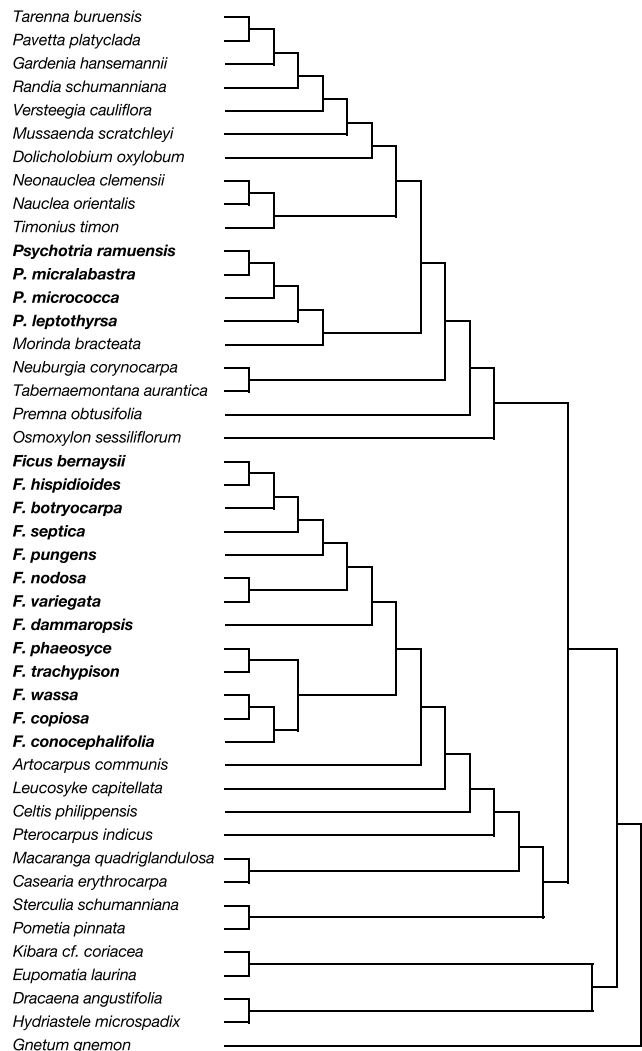


Figure 1 Phylogenetic relationships of host plants included in the study.

measures of host specificity that relied on counts of higher plant taxa (for example, genera or families). This approach can be misleading when taxonomic ranks are not commensurate with plant lineages. We examined the impacts of sampling bias and phylogenetic effects on estimates of host specificity by analysing the largest available data set of its kind. The leaf-chewing insect community on 51 plant species was characterized by using a sample

Table 1 Overlap between leaf-chewing communities from closely and distantly related host plants

Host plant	Herbivores	<i>r</i>	So (mean ± s.e.m.)
<i>Ficus</i> spp.	Coleoptera	–0.182	0.51 (±0.008)
	Lepidoptera	–0.274	0.52 (±0.010)
	Orthopteroids	–0.267	0.48 (±0.019)
	Total	–0.370	0.51 (±0.007)
			0.57 (±0.059)
<i>Psychotria</i> spp.	Coleoptera	0.58 (±0.020)	
	Lepidoptera	0.57 (±0.059)	
	Orthopteroids	0.54 (±0.076)	
	Total	0.57 (±0.029)	
			0.45 (±0.006)
Plant genera	Coleoptera	–0.237	
	Lepidoptera	–0.328	0.09 (±0.005)
	Orthopteroids	0.018	0.53 (±0.007)
	Total	–0.165	0.37 (±0.004)

r, Spearman correlation between the phylogenetic distance of plants and the overlap of their herbivore communities measured by the Sorensen index. Significant values (*P* < 0.05, Mantel test) are in bold; data for *Psychotria* were too limited for calculation. So, average value of the Sorensen index for all pairwise comparisons between communities from different hosts.

of 50,734 herbivores comprising 935 species in a tropical lowland forest in New Guinea.

An analysis of herbivore specificity compared the degree of faunal overlap between communities feeding on closely and distantly related plants (Fig. 1). Any two host species from within the genera *Ficus* and *Psychotria* shared ~50% of their herbivorous species (Fig. 2, Table 1). This overlap between herbivore communities decreased gradually with increasing phylogenetic distance between hosts. The same was true for each of the three major herbivorous groups, including moths and butterflies (Lepidoptera), beetles (Coleoptera) and grasshoppers and stick insects (orthopteroids). Of 210 *Ficus* and *Psychotria* herbivores with known host range, only one species of *Coenobius* (Chrysomelidae) feeding on *Ficus nodosa* was monophagous. Large overlap among neotropical herbivore communities on congeneric hosts was found also in *Passiflora*⁹ (but see ref. 10).

We postulate that monophagous herbivores are rare in tropical forests because large genera tend to dominate tropical floras on all spatial scales, and our data demonstrate that monophagy is scarce in such genera. For example, Rubiaceae, the world's largest family of predominantly tropical woody plants, includes 15 genera with at least 100 species, representing 42% of 10,200 species in the family¹¹. These genera also represent 50% of Rubiaceae in Costa Rica¹², 39% in New Guinea¹³ and 36–58% in four neotropical forest communities¹⁴. Furthermore, large genera feature prominently in most tropical plant families¹¹.

The overlap between herbivore communities from phylogenetically distant hosts was also high, although lower than in congeneric hosts (Fig. 2, Table 1). Communities of Coleoptera and orthopteroids maintained high overlap (more than 40%) even between distantly related hosts. Most species of Coleoptera (96%) fed on the foliage as adults only. The key determinant of the distribution for these species is probably the host range of their poorly known wood-boring or root-feeding larvae¹⁵. The overlap between orthopteroid communities did not even depend on phylogenetic distance between hosts, as many species had broad host ranges spanning multiple plant families. In contrast, communities of Lepidoptera feeding on plants from different genera and families were highly distinct, with a low overlap (less than 20%) that decreased steeply towards zero with increasing phylogenetic distance between hosts. The narrow host specificity of larval Lepidoptera in New Guinea is consistent with data from the neotropics^{3,16}.

Large (speciose) plant genera hosted more distinct leaf-chewing communities than small genera, as indicated by a negative correlation (Spearman $r = -0.38$, $P < 0.05$, $n = 29$, Independent Contrasts) between the size of the plant genus in New Guinea¹³ and the average overlap (S_o) of its herbivore community with communities feeding on other genera. The overall relationship was due to

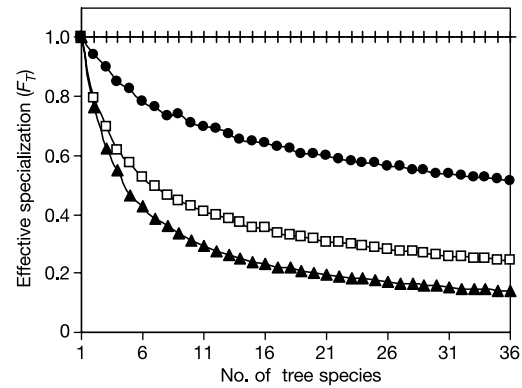


Figure 3 Dependence of effective specialization (F_T) on the number of host tree species studied. Samples from individual tree species, each from a different genus, were amalgamated in randomized sequence; the mean from 100 random sequences is given. Coleoptera (squares), Lepidoptera (circles), orthopteroids (triangles) and butterflies (crosses) are shown separately.

Coleoptera ($r = -0.45$, $P < 0.05$) and Lepidoptera ($r = -0.39$, $P < 0.05$); no correlation existed in the orthopteroids ($P > 0.5$). These findings contradict the hypothesis that taxonomically isolated plants from small genera support particularly distinct herbivore communities. We suggest that herbivores specialize more often on large plant genera than small ones because the former represent a large resource base, particularly because many herbivores have broad congeneric host ranges.

Our study showed that most herbivores feed on several closely related congeneric plant species. Many herbivores have even wider host ranges, as communities from allogeneric hosts shared on average 37% of all herbivorous species (Table 1). These results are at variance with the notion of extremely specialized interactions between herbivores and finely divided plant resources that provide distinct ecological niches and permit the coexistence of numerous herbivorous species in diverse tropical forests. Our findings suggest that alternative, non-equilibrium models of tropical diversity⁶ should be seriously considered.

There were on average $S_{av} = 32.9 \pm 1.7$ species of Coleoptera, 26.1 ± 2.4 species of Lepidoptera and 20.6 ± 0.9 species of orthopteroids feeding on a single host tree species. Their effective specialization¹⁷ was respectively $F_T = 0.24$ for Coleoptera, 0.51 for Lepidoptera and 0.14 for orthopteroids (Fig. 3) so that there were $S_{av}F_T = 7.9$ unique (effectively specialized) species of Coleoptera, 13.3 of Lepidoptera and 2.9 of orthopteroids per host tree species. These results are similar to the two to five unique species of

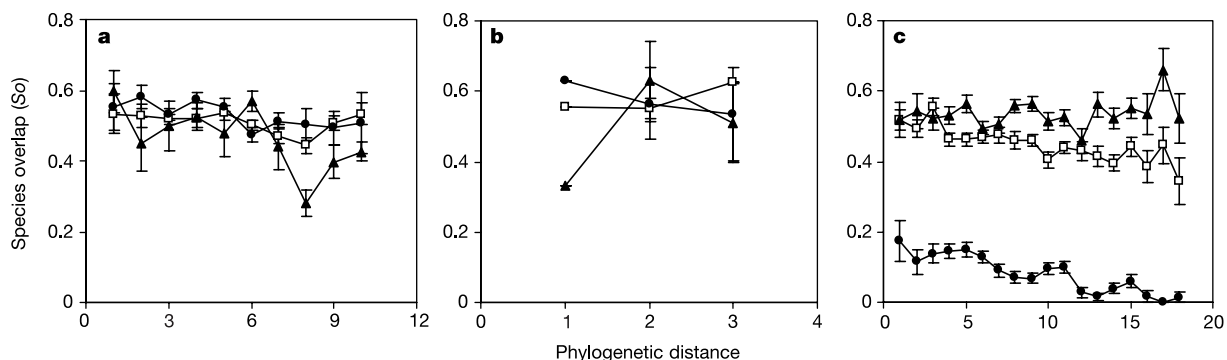


Figure 2 Overlap in species composition of their leaf-chewing communities and phylogenetic distance between *Ficus* spp. (a), *Psychotria* spp. (b) and plants from different genera (c). Values of Sorensen index S_o (means $\pm$ s.e.m.) are shown for all pairwise

comparisons between hosts separated by a particular phylogenetic distance. Coleoptera (squares), Lepidoptera (circles) and orthopteroids (triangles) are shown separately.

Table 2 Estimates of the global number of arthropod species

Variable	1	Coleoptera 2	3	Lepidoptera
A. No. of herbivore species per tree species	682*	32.9†	32.9†	30.8¶
B. Effective specialization	0.20*	0.24†	0.24†	0.51†
C. Correction for non-herbivorous species	1.20*	1.20*	1.20*	1.00†
D. Proportion of species from arthropods	0.40*	0.23‡	0.23‡	0.10#
E. Proportion of canopy fauna from total	0.66*	0.42‡	0.42‡	1.00†
F. No. of tropical tree species	50,000*	50,000*		
G. No. of plant genera in New Guinea			1,872§	1,872§
H. Proportion of New Guinea species from total			0.05	0.05
Global no. of arthropod species ($A \times B \times C \times F / (D \times E)$)	31,000,000	4,904,348		
Global no. of arthropod spp. ($(A \times B \times C \times G) / (D \times E \times H)$)			3,672,376	5,881,075

Sources: *, from ref. 1; †, from this study; ‡, from ref. 4; §, from ref. 13; ||, from ref. 19; ¶, our estimate of 26.1 externally feeding leaf-chewer species multiplied by 1.18 to account for herbivorous species from other guilds, based on ref. 30; #, from ref. 8. Coleoptera 1, estimate from ref. 1. Note that the Coleoptera 2 and Lepidoptera estimates are mutually independent because they are based on two entirely different sets of parameters.

Coleoptera per host tree reported from other tropical forests^{4,18} but at variance with the 136 unique species postulated by Erwin¹. Our estimates of effective specialization are robust, as they changed little with increasing number of tree species included in the analysis (Fig. 3). In contrast, S_{av} values were underestimated, as the species accumulation curves for individual tree species did not approach an asymptote, indicating that the total species richness of their herbivore assemblages had not been sampled.

These host specificity data, together with improved estimates of some other parameters⁴, permitted a revision of Erwin's¹ estimate of arthropod diversity from 31.0 to 4.9 million species (Table 2). Erwin extrapolated the species richness of a herbivore community from a single tree species to the total number of tropical tree species. We refined Erwin's method in recognition of the fact that plant genera, rather than species, are more appropriate units for estimating the diversity of tropical herbivore communities. The average number of herbivorous species per host species was multiplied by the 1,872 plant genera known from New Guinea¹³ to obtain an estimate of arthropod species in New Guinea. The New Guinean diversity was extrapolated to global, using an estimate¹⁹ that ~5% of the world diversity in several well-known groups of organisms, including flowering plants, various invertebrates and vertebrates, occur in New Guinea. The total arthropod diversity was estimated at 3.7 million and 5.9 million species from data on Coleoptera and Lepidoptera, respectively (Table 2). This method was verified by using butterflies, a well-known taxon with 959 described species in New Guinea²⁰ and 15,000–20,000 worldwide⁵. In our samples, butterflies had $S_{av} = 0.63$ species per host tree species and effective specialization $F_T = 1.0$, leading to an estimate of 1,179 species in New Guinea and 23,500 worldwide.

Our estimates of global diversity are similar to the 6.6 million (ref. 18) and 4.8 million (ref. 4) species of arthropods extrapolated from other communities of tropical beetles. Even more importantly, they are also in agreement with estimates of ~5 million species based on the analysis of regional faunas and the evaluation of museum collections by taxonomists^{5,7,8,21}, reconciling the disparate results of ecological and taxonomic approaches to the estimation of global species diversity. □

Methods

Insects

Leaf-chewing insects feeding on 51 woody species (Fig. 1 and the following species of Euphorbiaceae s.l.: *Breynia cernua*, *Endospermum labios*, *Homalanthus novoguineensis*, *Mallotus mollissimus*, *Melanolepis multiglandulosa* and *Pimelodendron amboinicum*) from all major lineages of flowering plants and all stages of rainforest secondary succession²² were studied at four study sites in a diverse lowland rainforest (>150 woody species per hectare) near Madang in Papua New Guinea in 1994–2000 (ref. 23). Insects were collected by hand from foliage of each plant species for at least 1 year. Collecting effort was 1,500 m² of foliage sampled and >1,000 tree inspections (a particular tree sampled at a particular time) per species. This sampling effort represented ~1,000 person-days of fieldwork. Each insect was provided with leaves of the plant species from which it was collected and only those that fed were retained in the analyses. Larvae were reared to adults whenever possible. All insects were assigned to morphospecies and later verified and identified by

specialists as far as possible (http://www.nmnh.si.edu/new_guinea/). Vouchers are deposited in the Bishop Museum, Honolulu, and the Smithsonian Institution, Washington.

Feeding

All feeding records limited to a single individual on a particular host species were excluded as poorly supported. Monophagous species were defined as those represented by at least 10 individuals, all feeding on a single plant species. Faunal overlap was examined in 45 host taxa with known phylogenetic relationships. Overlap was measured by the Sørensen coefficient $So = 2a / (2a + b + c)$, where a is the number of species common to the two compared samples, and b and c are the numbers of species present only in the first and second samples respectively. So quantifies the proportion of species in a sample that also occurs in the other sample, averaged over the two samples.

The average proportion of herbivorous species feeding on a particular host plant that was unique (effectively specialized¹⁷) to that plant was estimated as $F_T = S_{\eta} / H_T$, that is, as the ratio of the total number of herbivorous species found on all T hosts studied (S_{η}), divided by the number of host-plant records involving these hosts (H_T). The effective specialization and species richness of herbivores were estimated from communities feeding on 36 tree species, each representing a different genus.

Phylogenetic relationships

Phylogenetic distance between plants was estimated as the number of nodes in the phylogeny between these plant species. Separate correlation analyses of 13 *Ficus* spp., 4 *Psychotria* spp. and 30 plants from different genera were performed. *Ficus wassa* and *Psychotria micralabastra* represented their respective genera in the latter analysis. COMPARE 4.4 software was used to calculate independent contrasts.

Phylogenetic relationships of plant species were obtained from refs 24–28. All Rubiaceae species were sequenced for large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase gene (*rbcL*) and 30S ribosomal protein S16 gene (*rps16*). The data were analysed under parsimony with PAUP, version 4.0d64* (D. Swofford, unpublished). All relationships were highly supported and in agreement with earlier analyses of the family²⁹. Accession numbers from the EMBL/GenBank data bases for the *rps16* intron and the *rbcL* gene, respectively, follow (see Fig. 1 for full plant names): *G. hansemanii*, AJ320077, AJ318446; *M. citrifolia* chromoplast, AJ320078, AJ318448; *M. scratchleyi*, AJ320079, AJ318447; *N. orientalis*, AJ320080, AJ318449; *N. clemensiae*, AJ320081, AJ318450; *P. platyclada*, AJ320082, AJ318451; *P. leptothyrsa*, AJ320083, AJ318452; *P. micralabastra*, AJ320084, AJ318453; *P. micrococca*, AJ320085, AJ318454; *P. ramuensis*, AJ320086, AJ318455; *R. schumanniana*, AJ320087, AJ318456; *T. buruensis*, AJ320088, AJ318457; *T. timon*, AJ320089, AJ318458; *V. cauliflora*, AJ320090, AJ318459.

Received 6 September 2001; accepted 12 January 2002.

- Erwin, T. L. Tropical forests: their richness in Coleoptera and other species. *Coleopterist's Bull.* **36**, 74–75 (1982).
- Basset, Y. Host specificity of arboreal and free-living insect herbivores in rain forests. *Biol. J. Linn. Soc.* **47**, 115–133 (1992).
- Barone, J. A. Host-specificity of folivorous insects in a moist tropical forest. *J. Anim. Ecol.* **67**, 400–409 (1998).
- Ødegaard, F. How many species of arthropods? Erwin's estimate revised. *Biol. J. Linn. Soc.* **71**, 583–597 (2000).
- Stork, N. E. How many species are there? *Biodivers. Conserv.* **2**, 215–232 (1993).
- Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, 2001).
- May, R. M. in *Nature and Human Society: The Quest for a Sustainable World* (eds Raven, P. H. & Williams, T.) 30–45 (National Academy Press, Washington DC, 2000).
- Nielsen, E. S. & Mound, L. A. in *Nature and Human Society: The Quest for a Sustainable World* (eds Raven, P. H. & Williams, T.) 312–322 (National Academy Press, Washington DC, 2000).
- Thomas, C. D. Herbivore diets, herbivore colonization, and the escape hypothesis. *Ecology* **71**, 610–615 (1990).
- Marquis, R. J. in *Plant–Animal Interactions: Evolutionary Ecology in Tropical and Temperate Regions* (eds Price, P. W., Lewinsohn, T. M., Fernandes, G. W. & Benson, W. W.) 179–208 (Wiley, London, 1991).
- Mabberley, D. J. *The Plant Book* (Cambridge Univ. Press, Cambridge, 1987).
- Burger, W. & Taylor, C. M. Flora Costaricensis: Family #202 Rubiaceae. *Fieldiana Botany*, N.S. **33**, 1–333 (1993).

13. Höft, R. *Plants of New Guinea and the Solomon Islands. Dictionary of the Genera and Families of Flowering Plants and Ferns* (Wau Ecology Institute, Wau, 1992).
14. Gentry, A. H. (ed.) *Four Neotropical Forests* (Yale Univ. Press, New Haven, 1990).
15. Jolivet, P. & Hawkeswood, T. J. *Host-Plants of Chrysomelidae of the World* (Backhuys, Leiden, 1995).
16. Janzen, D. H. Ecological characterization of a Costa Rican dry forest caterpillar fauna. *Biotropica* **20**, 120–135 (1988).
17. May, R. M. How many species? *Phil. Trans. R. Soc. Lond. B* **330**, 293–304 (1990).
18. Basset, Y., Samuelson, G. A., Allison, A. & Miller, S. E. How many species of host-specific insects feed on a species of tropical tree? *Biol. J. Linn. Soc.* **59**, 201–216 (1996).
19. Sekhran, N. & Miller, S. E. (eds) *Papua New Guinea Country Study on Biological Diversity* (Papua New Guinea Dept of Environment & Conservation, Waigani, Papua New Guinea, 1996).
20. Parsons, M. *The Butterflies of Papua New Guinea* (Academic, London, 1999).
21. Miller, S. E., Novotny, V. & Basset, Y. in *Foundations of Tropical Biology: Key Papers and Commentaries* (eds Chazdon, R. L. & Whitmore, T. C.) (Univ. Chicago Press, Chicago, in the press).
22. Leps, J., Novotny, V. & Basset, Y. Habitat and successional optimum of plants and the composition of their leaf-chewing herbivores in Papua New Guinea. *J. Ecol.* **89**, 186–199 (2001).
23. Basset, Y., Novotny, V., Miller, S. E. & Pyle, R. Quantifying biodiversity: experience with parataxonomists and digital photography in Papua New Guinea and Guyana. *BioScience* **50**, 899–908 (2000).
24. Angiosperm Phylogeny Group An ordinal classification for the families of flowering plants. *Ann. Missouri Bot. Gard.* **85**, 531–553 (1998).
25. Weiblen, G. D. Phylogenetic relationships of functionally dioecious *Ficus* (Moraceae) based on ribosomal DNA sequences and morphology. *Am. J. Bot.* **87**, 1342–1357 (2000).
26. Backlund, M., Oxelman, B. & Bremer, B. Phylogenetic relationships within the Gentianales based on ndhF and rbcL sequences, with particular reference to the Loganiaceae. *Am. J. Bot.* **87**, 1029–1043 (2000).
27. Soltis, D. E. *et al.* Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Bot. J. Linn. Soc.* **133**, 381–461 (2000).
28. Qiu, Y.-L. *et al.* The earliest angiosperms. Evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**, 404–407 (1999).
29. Bremer, B. *et al.* More characters or more taxa for a robust phylogeny—case study from the coffee family (Rubiaceae). *Syst. Biol.* **48**, 413–435 (1999).
30. Powell, J. A., Mitter, C. & Farrell, B. in *Lepidoptera, Moths and Butterflies Vol. 1: Evolution, Systematics, and Biogeography* (ed. Kristensen, N. P.) 403–422 (Walter de Gruyter, Berlin, 1998).

Acknowledgements

We thank parataxonomists J. Auga, W. Boen, C. Dal, S. Hiuk, B. Isua, M. Kasbal, R. Kutil, M. Manumbor and K. Molem and also K. Darrow and N. Heidari for assistance; many insect collectors and taxonomists are acknowledged elsewhere. N. Stork, T. Lewinsohn, J. Zrzavy, J. Leps and A. Stewart commented on the manuscript. This work was supported by the National Science Foundation (USA), Christensen Fund (USA), Grant Agency of the Czech Republic, Czech Academy of Sciences, the Swedish Natural Science Research Council, Czech Ministry of Education, Otto Kinne Foundation, Darwin Initiative (UK), International Centre of Insect Physiology and Ecology (ICIPE) and Bishop Museum.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.N.
(e-mail: novotny@entu.cas.cz).

Developmental constraints versus flexibility in morphological evolution

Patricia Beldade, Kees Koops & Paul M. Brakefield

Institute of Evolutionary and Ecological Sciences, Leiden University,
PO Box 9516, 2300 RA Leiden, The Netherlands

Evolutionary developmental biology has encouraged a change of research emphasis from the sorting of phenotypic variation by natural selection to the production of that variation through development¹. Some morphologies are more readily generated than others, and developmental mechanisms can limit or channel evolutionary change². Such biases determine how readily populations are able to respond to selection³, and have been postulated to explain stasis in morphological evolution⁴ and unexplored morphologies⁵. There has been much discussion about evolutionary constraints^{6–8} but empirical data testing them directly are

sparse^{9,10}. The spectacular diversity in butterfly wing patterns¹¹ is suggestive of how little constrained morphological evolution can be. However, for wing patterns involving serial repeats of the same element, developmental properties suggest that some directions of evolutionary change might be restricted^{12,13}. Here we show that despite the developmental coupling between different eyespots in the butterfly *Bicyclus anynana*, there is great potential for independent changes. This flexibility is consistent with the diversity of wing patterns across species and argues for a dominant role of natural selection, rather than internal constraints, in shaping existing variation.

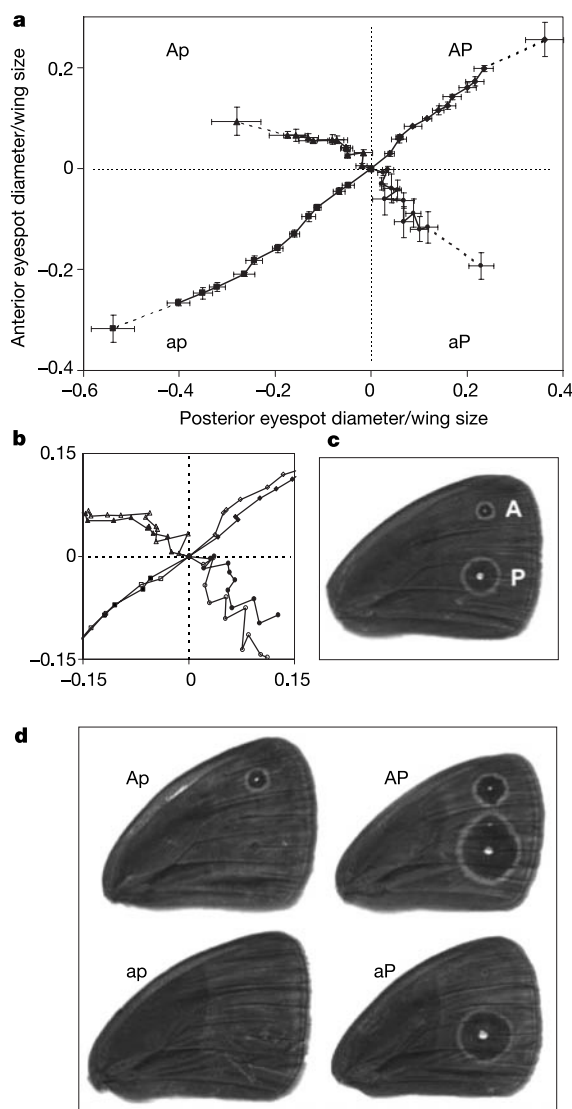


Figure 1 Response to artificial selection on the size of the dorsal forewing eyespots of *B. anynana*. **a**, Eyespot diameter/wing size relative to unselected control values are given for the different directions of selection. AP and ap are coupled directions; Ap and aP are uncoupling directions. Each point represents the mean ($\pm$ standard error) for the two replicate lines for each generation. Solid lines join points covering the first 11 consecutive generations, all starting from the same original population (centre of graphic, G0). Broken lines join the points for G11 and G25 phenotypes. **b**, Enlargement of the central area of **a** showing the behaviour of individual replicate lines (filled and open symbols). **c**, Typical dorsal surface of forewing of unselected female showing the anterior (A) and posterior (P) eyespots. **d**, Representative G25 phenotypes (most ap females have no eyespots and many AP females have extra, satellite eyespots). Responses in males were comparable to those in females.

The different pattern elements on butterfly wings, including the eyespots, are apparently generated by an extremely flexible developmental system that allows independent variation and evolution of these elements¹⁴. Nonetheless, pattern elements belonging to the same homologous series have frequently been found to be positively correlated within different species^{15,16}. Genetic covariances describe potential constraints arising from shared developmental pathways¹⁷ that might limit wing pattern evolution¹⁸. In *B. anynana*, eyespots are characterized by a conserved pattern of relative size and are formed by a common developmental mechanism. Each eyespot is centred around a group of organizing cells¹⁹ with a characteristic expression of several developmental genes^{20–22}. Furthermore, allelic combinations favoured by artificial selection for their effect on one eyespot generally affect other eyespots in the same direction^{16,23}. This coupling of eyespots might impose constraints on the evolution of wing patterns, at least in the short term, so we can predict that concerted changes will be more readily produced than opposing ones¹².

Artificial selection can be a powerful way of exploring the space of possible phenotypes^{2,24,25}. In this study we used this approach to examine how readily the pattern of relative size of the anterior and posterior eyespots on the dorsal forewing of *B. anynana* (Fig. 1c) can be changed on the basis of standing genetic variation present in a laboratory population. Despite the positive genetic correlation between eyespot sizes¹⁶, we obtained substantial responses to selection in different directions (Methods), both when the two eyespots were selected to change in a concerted or ‘coupled’ manner (both larger or both smaller, lines ‘AP’ and ‘ap’, respectively), and in opposite or ‘uncoupling’ directions (lines ‘Ap’ and ‘aP’) (Fig. 1). Directional selection was applied in two periods, from generations 0 to 11 and from G19 to G25 (with relaxed selection in between). Responses to selection were rapid, gradual and similar across replicate lines (Figs 1 and 2). All directions have produced butterflies with strikingly different ratios of eyespot size to wing size for both eyespots (Fig. 2), and by G25 almost all butterflies had extreme phenotypes not represented in the base population. These two domains of wing morphology in *B. anynana* have been able to respond independently to selection in a manner comparable to experiments with *Drosophila* wing compartments²⁵. Despite this autonomy, we have obtained strong correlated responses across wing surfaces with ventral eyespots showing substantial, albeit less extreme, changes in the different directions (results not shown).

The observation that the ‘uncoupling’ phenotypes are not as extreme as the ‘coupled’ phenotypes (Fig. 1) is not in itself evidence for constraints. Positive correlations between the size of the target eyespots ($r_{\text{Pearson}} = 0.52 \pm 0.02$ in G0 females²⁶) result in higher phenotypic variation along the coupled axis. Comparing changes in

eyespot size relative to cumulated selection differential across directions shows no clear evidence that response in the uncoupling lines is more difficult than in the coupled lines (Table 1; compare rates of response for the posterior eyespot). Nevertheless, there are differences in the behaviour of the two types of lines. For both periods of directional selection, the anterior eyespot responds more slowly in both Ap and aP lines relative to AP and ap (Table 1). In particular, the relative lack of response in Ap after G6 (Fig. 1) presumably reflects exhaustion of genetic variation that increases the anterior eyespot with no (or opposite) effect on the posterior (although alleles that increase both eyespots simultaneously are still present; see response to selection imposed on the anterior eyespot alone, A in Fig. 2). Furthermore, our results show wider differences between replicate lines for the uncoupling directions, and a clear contrast between an apparent step-like progression in the uncoupling directions (in each generation one eyespot responding more than the other) versus a more linear progression in the coupled directions (Fig. 1b). Similar ‘erratic responses’ have been reported for antagonistic selection applied to positively correlated traits in other organisms²⁷.

Despite these indications of tension for uncoupling changes, our results clearly demonstrate great potential for independent changes of eyespot size in *B. anynana*. This potential is apparently not explored in this species, perhaps owing to stabilizing selection on the pattern. We examined the potential fitness disadvantages of uncoupled phenotypes by monitoring changes under relaxed selection (G11–G19). Such disadvantages should result in more rapid declines in frequency of extreme Ap and aP relative to AP and ap phenotypes. However, all lines reverted gradually towards control values with no clear distinction between coupled and uncoupling directions (Fig. 2). Nonetheless, even if in laboratory conditions there are no substantial fitness disadvantages in having an uncoupled phenotype, these may exist in nature, where sexual selection and interspecific interactions are likely to be more important.

Overall, our results show great flexibility in the formation and

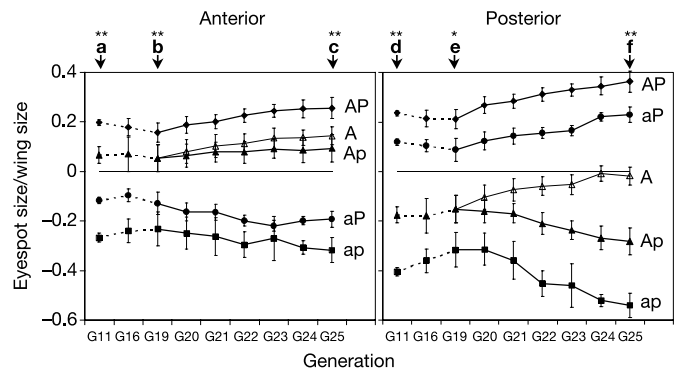


Figure 2 Response to relaxed selection (G11–G19) and additional directional selection (G19–G25). Average eyespot size/wing size values across replicate lines ($\pm$ standard error) are given relative to unselected control values (horizontal line) for the different directions. Directional selection produced butterflies with highly divergent phenotypes and after eight generations under relaxed selection, almost all lines reverted to values closer to the unselected control. Characters above arrows correspond to the following analysis of variance (ANOVA) *F* values (with degrees of freedom for factor/error) for the effect of direction on phenotype at G11, G19 and G25: **a**, 174.12 (4/6); **b**, 73.24 (4/5); **c**, 169.94 (5/6); **d**, 161.64 (4/6); **e**, 28.45 (4/5); and **f**, 119.16 (5/6). In all cases there were significant differences in eyespot size/wing size across directions (asterisk, $P = 0.001$; double asterisk, $P < 0.0005$). Tukey’s comparisons showed that all pairs of lines had significantly different eyespot size/wing sizes at a 5% error rate, with the following exceptions: **b**, Ap = UC; **c**, Ap = A; **e**, AP = aP = UC, ap = Ap; and **f**, AP = aP, UC = A.

Table 1 Rate of response to selection on *B. anynana* eyespot size

Direction	G0–G11		G19–G25	
	Anterior	Posterior	Anterior	Posterior
AP	0.331 ^a	0.301 ^c	0.281 ^d	0.308 ^e
ap	0.295 ^a	0.356	0.252 ^d	0.251 ^e
Ap	0.257 ^{a,b}	0.464	0.196 ^d	0.369 ^e
aP	0.237 ^b	0.253 ^c	0.126 ^d	0.355 ^e
A	NA	NA	0.247 ^d	NA
ANCOVA	9.84*** (3/40)	24.11*** (3/40)	3.18* (4/25)	3.96** (3/20)

Slopes of the regression lines of response to selection on cumulated selection differential are given for each target eyespot during each period of directional selection (see Methods). All regression coefficients are significantly different from zero with $P < 0.0005$, except those marked with † ($P < 0.025$). Realized heritabilities for the first period of directional selection range from 0.5 to 0.9. Analysis of covariance (ANCOVA) *F* values (and interaction/error degrees of freedom) are given for the interaction effect of direction with cumulated selection differential on eyespot size/wing size; *** $P < 0.0005$, ** $P = 0.023$, * $P = 0.03$. Superscript characters indicate pairs of values, for each eyespot, that are not significantly different under Tukey’s pairwise comparisons at 5% error rate. NA, not applicable.

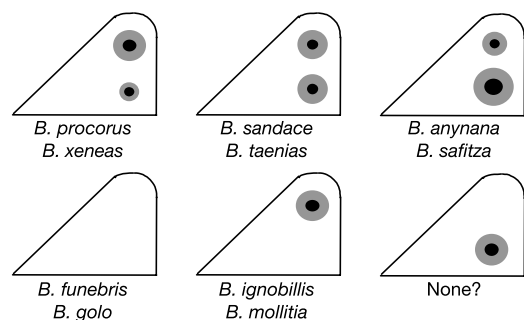


Figure 3 Diversity of eyespot patterns across the genus *Bicyclus*. Examples are shown of relative size phenotypes of the dorsal forewing eyespots of females from among the 80 or so *Bicyclus* species³⁰. Most patterns of relative eyespot size have been explored at a macroevolutionary scale, but that of a large posterior and no anterior eyespot appears not to be represented in any extant species³⁰. The pattern, however, eventually became frequent in both our aP selected lines of *B. anynana*. This suggests that its absence in the genus is explained by no history of selection in its favour rather than by an inability to generate it.

evolution of a morphological pattern for which developmental and genetic characterization had previously suggested the potential for constraints. This flexibility is consistent with the distribution of eyespot size phenotypes in the genus *Bicyclus* (Fig. 3). Standing genetic variation within a laboratory population of one species is sufficient to account for the production of all phenotypes found within the genus, and even one which is not explored in any extant species.

It has been proposed that the great flexibility in wing pattern formation in butterflies follows from the compartmentalization of individual pattern elements within vein-bounded wing regions^{14,28}. Our results provide experimental support for this proposal, and show that the developmental properties of eyespot formation are unlikely to constrain any process of adaptive radiation in the pattern of relative size of butterfly eyespots. Natural selection, together with population-level properties²⁹, rather than the generation of phenotypic variation, is likely to dominate in shaping the evolution of morphology that has led to the spectacular diversity of butterfly wing colour patterns. □

Methods

Target traits and directions of selection

Typically *B. anynana* butterflies have a small anterior and a large posterior eyespot on the dorsal surface of their forewings (Fig. 1c). We selected on the ratios between eyespot diameters (anterior, A , and posterior, P) and a measurement of wing size (W , the distance between two wing landmarks). Starting from a single outbred laboratory stock, we derived three groups of lines selecting on different combinations of eyespot/wing sizes: (1) selection for one eyespot to become larger and the other smaller, the 'uncoupling' directions (large anterior and small posterior eyespots, 'Ap', or small anterior and large posterior eyespots 'aP'), (2) selection on both eyespots to change in a concerted manner (that is, both larger, 'AP' or both smaller, 'ap'), the 'coupled' changes, and (3) the unselected controls (UC). We derived two replicate lines for each mode of directional selection and three replicate UC lines. Selection was done on the basis of an additive combination of the rank values of A/W (R_A) and P/W (R_P); $R_A + R_P$ for the coupled and $R_A - R_P$ for the uncoupling lines. The laboratory stock and the maintenance of the butterflies was described in previous studies^{16,23}.

Selection procedure

This experiment was divided into three consecutive phases: (1) 11 generations of directional selection of similar intensity for all lines (G0 to G11); (2) eight generations under relaxed selection (G11 to G19); and (3) six additional generations of directional selection with doubled intensity for the uncoupling lines (G19 to G25).

A total of 2,254 female butterflies were measured at G0 and 45 of these were randomly selected to lay eggs that produced the next generation of one UC replicate line. The remaining butterflies were randomly split into two groups from which the two sets of replicate lines for all other directions were derived (first the UCs and next the directional selection lines). In subsequent generations, between 150 and 200 females were measured per line. The selected females were put with about 50 random males and allowed to lay

eggs. To increase selection intensity, the number of parents was progressively reduced in the course of the experiment (no indication of inbreeding depression was observed). From G1 to G5 we selected 40 females per line; from G5 to G8, 35 females and from G8 to G11, 30. After G11 one UC line was discontinued and all other lines were maintained under relaxed selection with about 200 adult females reared per line per generation and 50–100 taken randomly as parents. After G19, the coupled lines and the UCs were maintained under a selection regime similar to that used in the first phase, while selection intensity for the uncoupling lines was doubled (330–450 females measured and 30 selected per line per generation). During this second period of directional selection we included another set of two replicate lines derived from each G19 Ap line and selected to increase the value of A/W with no selection on P/W (A lines).

Statistical analysis

To test for differences in phenotype between selection directions we performed analyses of variance (ANOVA)²⁶ on G11, G19 and G25 eyespot/wing phenotypes using the mean values of the two (or three for UCs) replicate lines for each direction (Fig. 2). When ANOVAs showed evidence of a significant effect of direction on phenotype, Tukey pairwise comparisons were performed.

Least-square regressions were fitted to the average points for eyespot size/wing size ratios (relative to UC values) on cumulated selection differential for each direction in each period of directional selection (Table 1). Average points across replicate lines were used because there were no significant differences between replicates (as tested with an analysis of covariance, ANCOVA²⁶). Realized heritabilities are estimated as twice the absolute values of the slopes of the regression lines (selection done on females only). ANCOVAs were used to compare the slopes of the regression lines using 'direction' as a fixed factor and 'cumulated selection differential' as a covariate²⁶. When these analyses showed a significant effect of the interaction between these two factors on eyespot size/wing size, Tukey pairwise comparisons were performed.

Received 11 October 2001; accepted 21 January 2002.

1. Stern, D. L. Perspective: Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000).
2. Maynard-Smith, J. *et al.* Developmental constraints and evolution. *Q. Rev. Biol.* **60**, 265–287 (1985).
3. Wagner, G. P. The influence of variation and development constraints on the rate of multivariate phenotypic evolution. *J. Evol. Biol.* **1**, 45–66 (1988).
4. Williamson, P. G. in *Rates of Evolution* (eds Campbell, K. S. W. & Day, M. F.) (Allen & Unwin, London, 1987).
5. Gould, S. J. A developmental constraint in *Cerion*, with comments on the definition and interpretation of constraint in evolution. *Evolution* **43**, 516–539 (1989).
6. Gould, S. J. & Lewontin, R. C. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**, 581–598 (1979).
7. Antonovics, J. & van Tienderen, P. H. Ontogenophyloconstraints? The chaos of constraint terminology. *Trends Ecol. Evol.* **6**, 166–168 (1991).
8. Pigliucci, M. & Kaplan, J. The fall and rise of Dr Pangloss: adaptationism and the Spandrels paper 20 years later. *Trends Ecol. Evol.* **15**, 66–70 (2000).
9. Travasano, M. Mongold, J. A., Bennett, A. F. & Lenski, R. E. Experimental tests of the role of adaptation, chance, and history in evolution. *Science* **267**, 87–90 (1995).
10. Teotónio, H. & Rose, M. R. Variability in the reversibility of evolution. *Nature* **408**, 463–466 (2000).
11. Nijhout, H. F. *The Development and Evolution of Butterfly Wing Patterns* (Smithsonian Inst. Press, Washington, 1991).
12. Brakefield, P. M. The evolution–development interface and advances with the eyespot patterns in *Bicyclus* butterflies. *Heredity* **80**, 265–272 (1998).
13. Brakefield, P. M. Structure of a character and the evolution of butterfly eyespot patterns. *J. Exp. Zool.* **291**, 93–104 (2001).
14. Nijhout, H. F. Symmetry systems and compartments in lepidopteran wings—the evolution of a patterning mechanism. *Development* (Suppl.) **225**–233 (1994).
15. Paulsen, S. & Nijhout, H. F. Phenotypic correlation structure among elements of the color pattern in *Precis coenia* (Lepidoptera, Nymphalidae). *Evolution* **47**, 593–618 (1993).
16. Monteiro, A. F., Brakefield, P. M. & French, V. The evolutionary genetics and developmental basis of wing pattern variation in the butterfly *Bicyclus anynana*. *Evolution* **48**, 1147–1157 (1994).
17. Cheverud, J. M. Quantitative genetics and developmental constraints on evolution by selection. *J. Theor. Biol.* **110**, 155–171 (1984).
18. Paulsen, S. M. Quantitative genetics of butterfly wing color patterns. *Dev. Genet.* **15**, 79–91 (1994).
19. French, V. & Brakefield, P. M. Eyespot development on butterfly wings: the focal signal. *Dev. Biol.* **168**, 112–123 (1995).
20. Brakefield, P. M. *et al.* Development, plasticity and evolution of butterfly eyespot patterns. *Nature* **384**, 236–242 (1996).
21. Brunetti, C. *et al.* The generation and diversification of butterfly eyespot color patterns. *Curr. Biol.* **11**, 1578–1585 (2001).
22. Beldade, P., Brakefield, P. M. & Long, A. D. Contribution of *Distal-less* to quantitative variation in butterfly eyespots. *Nature* **415**, 315–318 (2002).
23. Monteiro, A. F., Brakefield, P. M. & French, V. Butterfly eyespots: the genetics and development of the color rings. *Evolution* **51**, 1207–1216 (1997).
24. Scharloo, W. in *Organisational Constraints on the Dynamics of Evolution* (eds Vida, G. & Maynard-Smith, J.) 197–210 (Manchester Univ. Press, 1983).
25. Weber, K. E. How small are the smallest selectable domains of form? *Genetics* **130**, 345–353 (1992).
26. Zar, J. H. *Biostatistical Analysis* (Prentice Hall, Upper Saddle River, 1999).
27. Roff, D. A. *Evolutionary Quantitative Genetics* (Chapman & Hall, New York, 1999).
28. Nijhout, H. F. Independent development of homologous pattern elements in the wing patterns of butterflies. *Dev. Biol.* **108**, 146–151 (1985).
29. Barton, N. & Partridge, L. Limits to natural selection. *BioEssays* **22**, 1075–1084 (2000).
30. Condamine, M. *Monographie du Genre Bicyclus* (Lepidoptera, Satyridae) (Inst. Fond. Afr. Noire, Dakar, 1973).

Acknowledgements

We thank V. French, M. Matos, H. Metz, H. Teotónio, G. Wagner and B. Zwaan for helpful discussions about the experimental design, M. Lavrijsen, E. Schlatmann, B. de Winter and N. Wurzer for cultivating maize for hungry larvae, and V. French and H. Teotónio for comments on the manuscript. This work was supported by the Portuguese Foundation for Science and Technology under the Gulbenkian PhD Program (P.B.) and the Human Frontiers Science Program (P.M.B.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.B. (e-mail: pbeldade@rulsfb.leidenuniv.nl) or P.M.B. (e-mail: brakefield@rulsfb.leidenuniv.nl).

Association of dwarfism and floral induction with a grape 'green revolution' mutation

Paul K. Boss* & Mark R. Thomas

CSIRO Plant Industry and Cooperative Research Centre for Viticulture, PO Box 350, Glen Osmond, SA 5064, Australia

The transition from vegetative to reproductive growth is an essential process in the life cycle of plants. Plant floral induction pathways respond to both environmental and endogenous cues and much has been learnt about these genetic pathways by studying mutants of *Arabidopsis*^{1,2}. Gibberellins (GAs) are plant growth regulators important in many aspects of plant growth and in *Arabidopsis* they promote flowering^{3–5}. Here we provide genetic evidence that GAs inhibit flowering in grapevine. A grapevine dwarf mutant derived from the L1 cell layer of the champagne cultivar Pinot Meunier produces inflorescences along the length of the shoot where tendrils are normally formed. The mutated gene associated with the phenotype is a homologue of the wheat 'green revolution' gene *Reduced height-1* (ref. 6) and the *Arabidopsis* gene *GA insensitive* (*GAI*)⁷. The conversion of tendrils to inflorescences in the mutant demonstrates that the grapevine tendril is a modified inflorescence inhibited from completing floral development by GAs.

Grapevine (*Vitis* sp.) is one of the world's major perennial horticultural crops. It is a vine, and under natural conditions tendrils are used to support a tree-climbing habit to reach high sunlight levels for flowering⁸. A small number of *Vitis vinifera* cultivars dominate wine production in the world owing to their reputation for producing premium quality wine, and in France the Champagne region has become famous for its sparkling wine. Pinot Meunier, Pinot noir and Chardonnay are the only three cultivars authorized to be grown for champagne production; together the black berry cultivars, Pinot Meunier and Pinot noir, represent 74% of the planted vines. Pinot Meunier is a cultivar of ancient origins and has long been considered a periclinal mutant of Pinot noir. It is distinguished from Pinot noir in having tomentose (densely covered with trichomes) shoot tips and expanding leaves^{9,10}. All grapevine cultivar propagation is vegetative, and novel phenotypes, like that of Pinot Meunier, arise by somatic mutation.

The apical meristem of the grapevine shoot is organized into two distinct layers designated L1 (outermost) and L2 (ref. 11). Plants have been regenerated from the L1 and L2 cell layers of Pinot

Meunier by passage through somatic embryogenesis, and whereas those from the L2 cell layer were phenotypically indistinguishable from Pinot noir, the plants regenerated from the L1 cell layer displayed the tomentose phenotype of Pinot Meunier and were dwarfed¹². When grown under glasshouse conditions favourable for floral induction, the L1 dwarf plants produced inflorescences and bunches along the length of the shoots (Fig. 1a, c) where the L2 plants (and Pinot Meunier) had a normal phenotype and produced tendrils (Fig. 1b, d). Inflorescences and tendrils in grapevines are derived from meristematic structures called uncommitted primordia (Fig. 1e), which develop from shoot meristems and are found opposite two of every three leaves¹³. Uncommitted primordia formed on actively growing shoots develop into tendrils (Fig. 1b, f), whereas those in latent buds develop into inflorescences. Latent buds are formed during spring and summer and experience a winter dormancy before bud burst and flowering (Fig. 1d). In the L1 dwarf plants this process is not necessary and uncommitted primordia differentiate into inflorescences on actively growing shoots (Fig. 1a, g).

The dwarf stature of the L1 plants was consistent with altered levels of GAs or an altered response to GAs. The application of GAs and inhibitors of GA biosynthesis has been shown to modify grapevine tendril and inflorescence development^{14–16}. We concluded, on the basis of the following, that the L1 plants had an altered GA response and that this is associated with a mutated gene similar to the *Arabidopsis* gene *GAI*^{7,17}, a negative regulator of GA response. First, the L1 plants did not respond when GA was applied, indicating that it was not a GA-deficient dwarf. Second, the L1 mutant accumulated fourfold more GA₁ and 12-fold more GA₄ in leaves than the L2 plant (data not shown). Elevated levels of

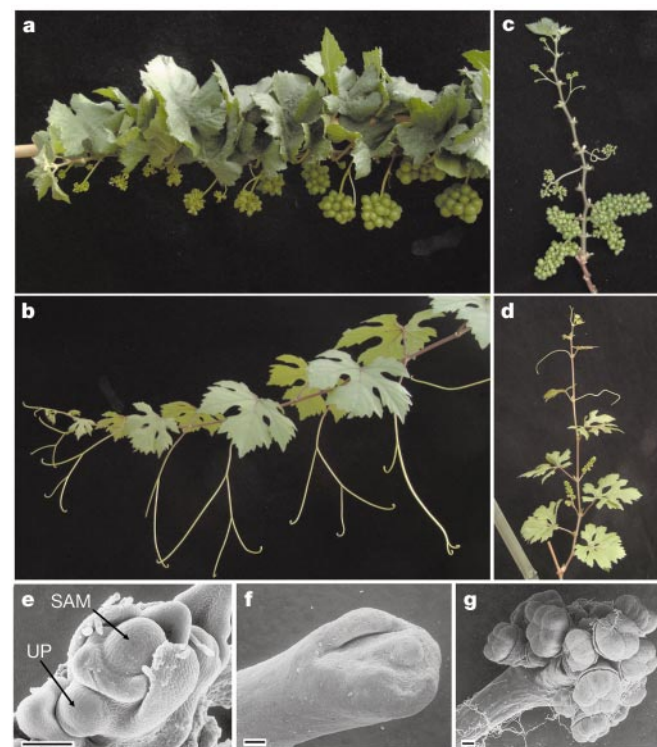


Figure 1 The L1 plant produces inflorescences instead of tendrils. **a**, Main shoot of an L1 plant. **b**, Main shoot of an L2 plant. **c**, Shoot from a latent bud of an L1 plant (leaves removed). **d**, Shoot from a latent bud of an L2 plant. **e**, Scanning electron micrograph of a shoot meristem from a wild-type latent bud showing an uncommitted primordium (UP) that has separated from the shoot apical meristem (SAM). **f**, Scanning electron micrograph of a tendril tip from an L2 plant. **g**, Scanning electron micrograph showing flowers at the tip of a tendril-like structure from an L1 plant. Scale bar in **e–g**, 100 μ m.

* Present address: John Innes Centre, Norwich Research Park, Norwich, Norfolk NR4 7UH, UK.

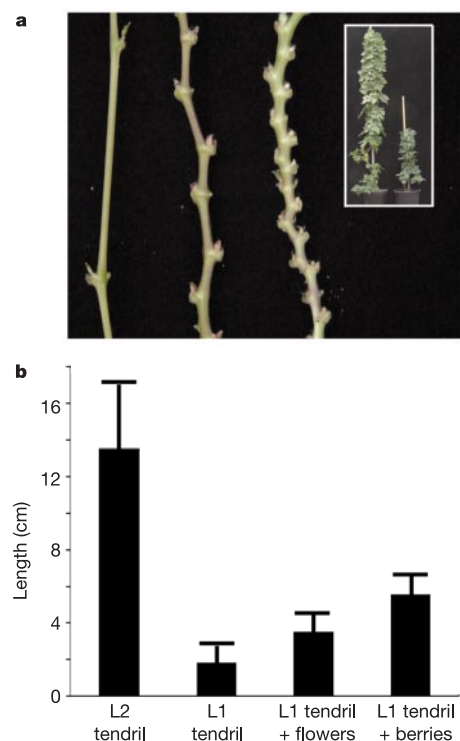


Figure 2 Mutant phenotypes. **a**, The stems of tall (left), semi-dwarf (middle) and dwarf (right) plants obtained from the L1 × L1 cross (leaves and bunches removed). The inset shows a semi-dwarf plant (left) and a dwarf plant (right) 3 months after germination. **b**, Tendril length on L2 plants (L2 tendril; *n* = 23), immature tendril-like structures on L1 plants (L1 tendril; *n* = 35), L1 tendril/inflorescence structures with terminal flowers (L1 tendril + flowers; *n* = 22) or L1 tendril/inflorescence structures with berries (L1 tendril + berries; *n* = 26). Results are means ± s.e.m.

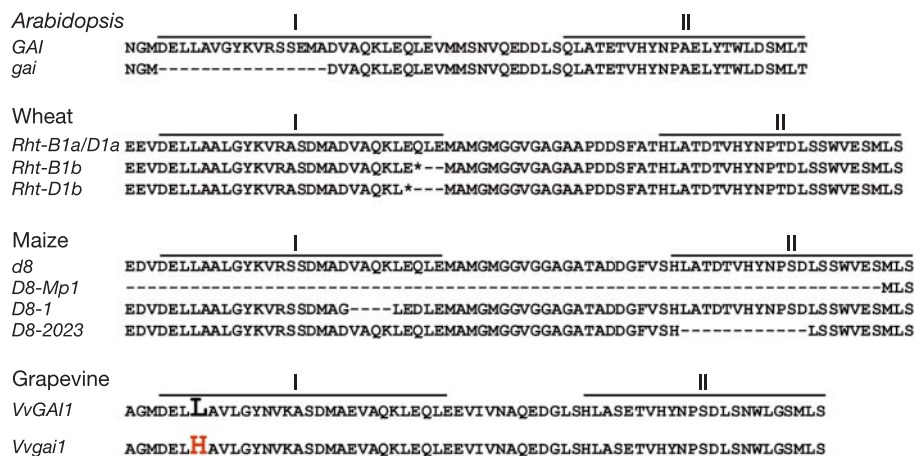


Figure 3 The *Vvgai1* point mutation results in an amino-acid substitution. The DELLA domains of the predicted proteins of wild-type alleles of the *Arabidopsis* *GAI* and homologues from maize (*d8*), wheat (*Rht-B1a*, *Rht-D1a*) and grapevine (*VvGAI1*) are compared to their mutant alleles^{6,7} below. Highly conserved regions of this domain,

named I and II by Peng *et al.*⁶, are shown. Deletions are represented by the dashed line and the position of translation stop codons are represented by an asterisk. A nucleotide substitution results in a leucine to histidine substitution in the I region of the DELLA domain of *VvGAI1*.

Table 1 *VvGAI1* genotype analysis of the progeny from crosses with the L1 dwarf

Cross	Genotype	Phenotype		
		Plant stature	Leaf-opposed structures	Shoot tips
L1 dwarf × L1 dwarf	<i>VvGAI1/VvGAI1</i> (29)	Tall	Tendrils	Non-tomentose
	<i>VvGAI1/Vvgai1</i> (57)	Semi-dwarfed	Flowers	Tomentose
	<i>Vvgai1/Vvgai1</i> (6)	Dwarfed	Flowers	Tomentose
<i>Vitis riparia</i> × L1 dwarf	<i>VvGAI1/VvGAI1</i> (23)	Tall	Tendrils	Non-tomentose
	<i>VvGAI1/Vvgai1</i> (5)	Semi-dwarfed	Flowers	Tomentose

The numbers of plants genotyped by using dCAPS are shown in parentheses.

The L1 plants were self-fertilized or crossed to *Vitis riparia* and the progeny were genotyped by using analyses with the dCAPS finder program²⁰. All of the tall progeny were homozygous wild-type (*VvGAI1/VvGAI1*) (Table 1). The dwarf progeny were either heterozygous (*VvGAI1/Vvgail1*) or homozygous (*Vvgail1/Vvgail1*) for the point mutation. The *Vvgail1/Vvgail1* plants were smaller than the heterozygotes and confirmed the semidominant nature of the mutation (Fig. 2a). Both dwarf genotypes had tomentose shoot tips and produced inflorescences from uncommitted primordia on actively growing shoots, whereas tall progeny were not tomentose and produced only tendrils. These data show that the dwarf, tomentose and prolific flowering phenotypes segregate with the mutated *Vvgail1* allele.

The change in GA response of the mutant also had a significant effect on seed dormancy and distorted the ratios of genotypes recovered from the crosses (Table 1). With the use of a typical method²¹, the seed germination rate was low, and only wild-type plants were recovered. When the seed was treated with GA₃ and scarified, higher germination rates and dwarf progeny were obtained. Without this treatment, seed containing the mutation remained dormant under normal germination conditions for more than a year.

Although most uncommitted primordia on *Vvgail1*-containing plants developed into inflorescences, some did not and resembled immature tendrils. Long tendrils did not form, indicating that a normal GA response is required for complete development (Fig. 2b). Conversion of the tendril to an inflorescence by the development of flowers (Fig. 1g), and later berries, stimulates the elongation of what has now become the rachis (Fig. 2b). This suggests that floral meristems and fruit have a role in determining inflorescence size and structure that is distinct from GA signalling via *VvGAI1*.

Gibberellin sprays are regularly used in the vineyard to increase the berry size of seedless grapevine varieties artificially. Surprisingly, the weight and volume of mature berries from L1 plants were not smaller than those of L2 or Pinot Meunier berries (data not shown). *VvGAI1* expression was not detected in berries (Supplementary Information, Fig. S4). This demonstrates that *VvGAI1* has no role in berry development and explains why fruit size is not reduced in the mutant.

Pinot Meunier is a very old cultivar and has been known since the 1500s (ref. 9). The existence of the *Vvgail1* mutation in Pinot Meunier pre-dates the introduction of mutant dwarfing alleles into cereals during the 'green revolution' of the 1960s and 1970s (ref. 6). The mutation seems to be cell-autonomous (as is suggested for *gai* (ref. 22), with only the tomentose phenotype being observed in the Pinot Meunier chimaera and no obvious effect on flowering or plant stature. The full effect of the mutation in grapevine was revealed only when the L1 and L2 layers were separated. In the vineyard, intensive management practices are required to control vine vigour and promote flowering and fruit development. The L1 mutant seems to possess desirable agronomic characteristics of both reduced vigour and increased fruiting potential.

Flowering in *Arabidopsis* involves a network of floral induction pathways (autonomous, photoperiod, vernalization and GA) that act redundantly to ensure that flowering occurs under appropriate conditions²³. Grapevine floral induction in latent buds is stimulated by high light intensity and high temperature rather than by day-length and vernalization²⁴. The association of the *Vvgail1* mutation with the marked effect on grapevine flowering indicates a major role for GAs in floral inhibition. Reproductive growth in the shoot is not totally inhibited by GAs because uncommitted primordia are still made, indicating that GA signalling through *VvGAI1* acts later by inhibiting the production of floral meristems and creates a new organ—a tendril. In the mutant, floral meristems are produced and normal tendrils are absent.

Vitis vinifera has evolved under very different conditions from those of *Arabidopsis*. Its natural habitat is forest, where it has

adapted to climbing trees to move from a shaded environment below the canopy to full sunlight above. It seems that GAs not only promote stem internode elongation to escape the shaded environment but also convert inflorescences into tendrils for enhanced climbing ability. Our findings suggest that whereas the transmission of the GA signal in the floral genetic pathway promotes flowering in *Arabidopsis*^{3–5}, it inhibits floral meristem production in grapevine. □

Methods

Plant material and seed germination

Grapevines were grown in commercial potting mix in a glasshouse with temperature and light/dark cycles of 16 h at 30 °C and 8 h at 25 °C. In winter months, the light duration was extended to 16 h through the use of artificial lighting. The Pinot Meunier L1 dwarf and the L2 vines with normal growth were regenerated from embryogenic callus as described¹². Self-pollinated L1 dwarf seed and seed obtained from outcrossing the L1 dwarf with *Vitis riparia* were cold-treated for at least 3 weeks before scarification and germination in light on moistened filter paper. Seedlings were transferred to soil and slowly acclimatized to glasshouse conditions.

PCR amplification of *GAI* homologues

Degenerate oligonucleotide primers were designed to internal conserved regions of known *GAI* homologues obtained from GenBank databases. Initial cloning was performed with Shiraz cDNA synthesized from immature inflorescence total RNA that was isolated as described²⁵. After the cloning of an internal *GAI*-like fragment, 5' and 3' rapid amplification of cloned ends (RACE) reactions²⁶ were used to obtain the full-length sequence for *VvGAI1*. Primers were used to isolate full-length *VvGAI1* alleles from Pinot Meunier, L1 and L2 plants. Sequences were compared from two independent amplifications from DNA and one amplification from cDNA to confirm differences.

DNA extraction and dCAPS analysis

DNA from each grapevine was extracted from leaves with a method described previously²⁷. An oligonucleotide primer was designed that would distinguish between the two species of *VvGAI1* sequences found in the Pinot Meunier L1 dwarf cDNA by using the dCAPS finder program²⁰. The primer introduced a *XhoI* site at the 3' end of the PCR product if the wild-type sequence was the template. After PCR amplification and digestion with *XhoI*, the presence of either *VvGAI1* allele could be determined by electrophoretic separation of the PCR products on a 2.5% agarose gel.

Received 4 January; accepted 14 February 2002.

- Hempel, F. D., Welch, D. R. & Feldman, L. J. Floral induction and determination: where is flowering controlled? *Trends Plant Sci.* **5**, 17–21 (2000).
- Araki, T. Transition from vegetative to reproductive phase. *Curr. Opin. Plant Biol.* **4**, 63–68 (2001).
- Blazquez, M. A., Green, R., Nilsson, O., Sussman, M. R. & Weigel, D. Gibberellins promote flowering of *Arabidopsis* by activating the *LEAFY* promoter. *Plant Cell* **10**, 791–800 (1998).
- Blazquez, M. A. & Weigel, D. Integration of floral inductive signals in *Arabidopsis*. *Nature* **404**, 889–892 (2000).
- Langridge, J. Effect of day-length and gibberellic acid on the flowering of *Arabidopsis*. *Nature* **180**, 36–37 (1957).
- Peng, J. et al. 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* **400**, 256–261 (1999).
- Peng, J. et al. The *Arabidopsis* *GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes Dev.* **11**, 3194–3205 (1997).
- Rives, M. Vigour, pruning, cropping in the grapevine (*Vitis vinifera* L.). I. A literature review. *Agronomie* **20**, 79–91 (2000).
- Viala, P. & Vermorel, V. *Ampélographie, Tome II* (Masson, Paris, 1901).
- Skene, K. G. M. & Barlass, M. Studies on the fragmented shoot apex of grapevine. IV. Separation of phenotypes in a periclinal chimera *in vitro*. *J. Exp. Bot.* **34**, 1271–1280 (1983).
- Thompson, M. M. & Olmo, H. P. Cytohistological studies of cytochimeric and tetraploid grapes. *Am. J. Bot.* **50**, 901–906 (1963).
- Franks, T., Botta, R. & Thomas, M. R. Chimerism in grapevines: implications for cultivar identity, ancestry and genetic improvement. *Theor. Appl. Genet.* **104**, 192–199 (2002).
- Pratt, C. Reproductive anatomy in cultivated grapes: a review. *Am. J. Enol. Vitic.* **22**, 92–109 (1971).
- Mullins, M. G. Regulation of inflorescence growth in cuttings of the grapevine (*Vitis vinifera* L.). *J. Exp. Bot.* **19**, 532–543 (1968).
- Utsunomiya, N., Sugiura, A. & Tomana, T. Effect of CCC and pinching on inflorescence induction and development on lateral shoots in grapevine. *J. Jpn. Soc. Hort. Sci.* **47**, 151–157 (1978).
- Boss, P. K. & Thomas, M. R. Tendrils, inflorescences and fruitfulness: A molecular perspective. *Aust. J. Grape Wine Res.* **6**, 168–174 (2000).
- Koornneef, M. et al. A gibberellin insensitive mutant of *Arabidopsis thaliana*. *Physiol. Plant.* **65**, 33–39 (1985).
- Peng, J. et al. Extragenic suppressors of the *Arabidopsis* *gai* mutation alter the dose-response relationship of diverse gibberellin responses. *Plant Physiol.* **119**, 1199–1207 (1999).
- Pysh, L. D., Wysocka-Diller, J. W., Camilleri, C., Bouchez, D. & Benfey, P. N. The GRAS gene family in *Arabidopsis*: Sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *Plant J.* **18**, 111–119 (1999).
- Neff, M. M., Neff, J. D., Chory, J. & Pepper, A. E. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: Experimental applications in *Arabidopsis thaliana* genetics. *Plant J.* **14**, 387–392 (1998).

21. Ellis, R. H., Hong, T. D. & Roberts, E. H. A note on the development of a practical procedure for promoting the germination of dormant seed of grape (*Vitis* spp.). *Vitis* **22**, 211–219 (1983).
22. Peng, J. & Harberd, N. P. Transposon-associated somatic *gai*-loss sectors in *Arabidopsis*. *Plant Sci.* **130**, 181–188 (1997).
23. Simpson, G. G., Gendall, A. R. & Dean, C. When to switch to flowering. *Annu. Rev. Cell Dev. Biol.* **15**, 519–550 (1999).
24. Buttrose, M. S. Climatic factors and fruitfulness in grapevines. *Hort. Abstr.* **44**, 319–326 (1974).
25. Boss, P. K., Davies, C. & Robinson, S. P. Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L. cv. Shiraz grape berries and the implications for pathway regulation. *Plant Physiol.* **111**, 1059–1066 (1996).
26. Frohman, M. A., Dush, M. K. & Martin, G. R. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl Acad. Sci. USA* **85**, 8998–9002 (1988).
27. Thomas, M. R., Matsumoto, S., Cain, P. & Scott, N. S. Repetitive DNA of grapevine: classes present and sequences suitable for cultivar identification. *Theor. Appl. Genet.* **86**, 173–180 (1993).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank R. King and A. Poole for the GA analyses by GC–MS; T. Franks for plant material; S. McClure for help with the scanning electron microscopy; and D. MacKenzie and C. Reich for technical assistance. This work was supported in part by the Grape and Wine Research and Development Corporation, Dried Fruits Research and Development Council and the Cooperative Research Centre for Viticulture.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.R.T. (e-mail: mark.r.thomas@csiro.au). The GenBank library accession number for *VvGAI1* is AF378125.

Emi1 is required for cytotstatic factor arrest in vertebrate eggs

Julie D. R. Reimann & Peter K. Jackson

Departments of Pathology, Microbiology and Immunology, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California 94305-5324, USA

Vertebrate eggs are arrested at metaphase of meiosis II with stable cyclin B and high cyclin B/Cdc2 kinase activity. The ability of the anaphase-promoting complex/cyclosome (APC), an E3 ubiquitin ligase, to trigger cyclin B destruction and metaphase exit is blocked in eggs by the activity of cytotstatic factor (CSF) (reviewed in ref. 1). CSF was defined as an activity in mature oocytes that caused mitotic arrest when injected into dividing embryos². Fertilization causes a transient increase in cytoplasmic calcium concentration leading to CSF inactivation, APC activation, cyclin B destruction and mitotic exit³. The APC activator Cdc20 is required for APC activation after fertilization^{4,5}. We show here that the APC^{Cdc20} inhibitor Emi1 (ref. 6) is necessary and sufficient to inhibit the APC and to prevent mitotic exit in CSF-arrested eggs. CSF extracts immunodepleted of Emi1 degrade cyclin B, and exit from mitosis prematurely in the absence of calcium. Addition of Emi1 to these Emi1-depleted extracts blocks premature inactivation of the CSF-arrested state. Emi1 is required to arrest unfertilized eggs at metaphase of meiosis II and seems to be the long-sought mediator of CSF activity.

CSF prevents parthenogenetic activation while animal eggs await fertilization. The molecular composition of CSF is unknown, although the MAP-kinase (MAPK) pathway is required to establish the CSF state in maturing oocytes. Specifically, the proto-oncogene Mos, a MAPK kinase kinase, is involved in the activation of

maturation promoting factor (cyclin B–cdc2 kinase complex) in meiosis I (MI) and during the transition from MI to meiosis II (MII) in *Xenopus* oocytes^{7,8}. Unfertilized oocytes from *Mos*^{−/−} mice undergo parthenogenetic activation^{9,10}. The Mos/MAPK pathway also prevents DNA replication between MI and MII—thereby allowing meiotic reduction in chromosome number—by prematurely reactivating cyclin B/Cdc2 after metaphase^{11,12}. However, the MAPK pathway is not sufficient for MII, because injection of the Mos protein is not sufficient to prevent mitotic exit after anaphase I in the presence of protein synthesis inhibitors^{8,11}, indicating that another factor is required to progress to metaphase of MII. The protein kinases MAPK kinase (MEK), MAPK and p90^{rsk} are essential downstream mediators of Mos activity^{13–16}. However, although Mos, MAPK and p90^{rsk} each cause a CSF-like metaphase arrest when injected into dividing embryos, neither Rsk depletion nor MAPK inactivation in CSF-arrested egg extracts releases CSF arrest¹³ (see below). Thus, the MAPK pathway is required to establish, but not to maintain, CSF arrest in MII.

Fertilization *in vivo* or the addition of calcium to extracts from unfertilized *Xenopus* eggs (called CSF extracts) activates APC, the ubiquitin ligase that mediates cyclin B destruction and mitotic exit^{3,17,18}. The APC activator Cdc20 is required for the calcium-induced exit from CSF arrest, indicating that APC^{Cdc20} is a target of the CSF inhibitory mechanism^{4,5}. Cdc20 is also a target of the spindle-assembly checkpoint protein Mad2, which inhibits APC activation by Cdc20 in response to spindle damage in the mitotic cell cycle¹⁹. Although Mad2 is present in CSF-arrested eggs²⁰, Mad2 and other spindle-assembly checkpoint proteins are not required for CSF arrest^{4,20–22}.

We recently identified Emi1 as an APC^{Cdc20} inhibitor in the *Xenopus* embryo⁶. In the mitotic cell cycle, Emi1 accumulates before mitosis and binds Cdc20 to inhibit its ability to activate the APC, thus allowing cyclin B to accumulate. In mitosis, Emi1 is ubiquitinated and destroyed independently of the APC. The microinjection of Emi1 into cleaving embryos, like that of Mos²³, constitutively active MAPK¹⁵ or constitutively active p90^{rsk} (ref. 14), causes a CSF-like mitotic arrest⁶. Given these results and the presence of Emi1 in the unfertilized egg, we were interested in testing whether Emi1 is a component of CSF.

We first examined whether excess Emi1 protein was sufficient to prevent release from the MII metaphase arrest in the presence of calcium. Cyclin B and Mos are stable and MAPK is active in extracts prepared from metaphase of MII-arrested eggs (CSF extracts). In CSF extracts treated with control protein, cyclin B is destroyed, Mos is inactivated and destroyed, and MAPK is inactivated, after calcium addition. The addition of purified maltose-binding protein (MBP)–Emi1 protein (1 μM, compared with ~300 nM endogenous Emi1 (ref. 6)) to CSF extracts prevented the calcium-induced destruction of cyclin B and Mos and the inactivation of MAPK (Fig. 1a). Although Mos is not an APC substrate, Mos is stabilized when the APC is inhibited because Mos destruction requires cyclin B/cdc2 inactivation²⁴. Examination of DNA morphology revealed that control extracts exited from mitosis by 15 min, whereas Emi1-treated extracts remained arrested in metaphase after more than 90 min (not shown). Inhibiting the APC by adding a destruction box (D-box) peptide also blocked CSF release (Fig. 1b). Thus, the APC is required for release from CSF arrest.

Mos inactivation and destruction occur later than cyclin B destruction after either egg activation or the addition of calcium to CSF extracts (Fig. 1a)²⁵. Incubation of CSF extracts with Mos protein before calcium addition did not prevent cyclin B destruction even though MAPK remained activated (Fig. 1c), further indicating that Mos inactivation is not required for exit from CSF arrest. Addition of the MEK inhibitor U0126 to CSF extracts led to the inactivation of MAPK and p90^{rsk} (ref. 12) but did not release extracts from the CSF state (Fig. 1c), which is consistent with previously reported results on depletion of the MEK target p90^{rsk}

from CSF extracts¹³. These data indicate strongly that the MAPK pathway is not required for maintenance of CSF arrest. To test whether Emi1 requires the MAPK pathway for inhibition of calcium-induced CSF release, we incubated CSF extracts with U0126 and found that Emi1 addition still prevented the destruction of cyclin B and Mos and exit from mitosis in the presence of calcium (Fig. 1c). Thus, Emi1's APC inhibitory activity does not require the MAPK pathway.

Calmodulin-dependent protein kinase II (CaMKII) is required for release of MII from metaphase, and constitutively active CaMKII is sufficient to trigger cyclin B destruction and mitotic exit without fertilization or the addition of calcium²⁶. Thus, the major calcium-sensitive mediator of CSF release is CaMKII. We found that constitutively active CaMKII did not trigger cyclin B destruction, Mos destruction or mitotic exit in CSF extracts in the presence of excess Emi1 (Fig. 1d), indicating that Emi1 acts downstream of CaMKII. These results are consistent with the observation that Cdc20 depletion also prevents constitutively active CaMKII-induced release from CSF arrest⁴. Additionally, much like Cdc20 depletion, excess Emi1 can block okadaic-acid-induced cyclin B destruction in CSF extracts in the absence of calcium (Supplementary Information, Fig. S1).

We next examined whether the Emi1 protein accumulates in maturing oocytes in sufficient time to account for CSF arrest. Emi1 is first expressed to significant levels in stage VI G2/prophase oocytes (Fig. 2a). After the addition of progesterone, Emi1 levels increased by germinal vesicle breakdown, similar to Cdc20 levels (Fig. 2a)²⁷. Emi1 levels and phosphorylation state did not seem to fluctuate significantly during oocyte maturation, whereas Cdc20 protein levels did (Fig. 2a)²⁷. In mature oocytes, Emi1 is present in sufficient amounts (~300 nM) to inhibit the endogenous pool of

Cdc20 (~40–100 nM (ref. 4)).

To test whether Emi1 is required for the maintenance of the MII metaphase arrest, we examined whether Emi1 depletion caused exit from CSF arrest. Incubation of CSF extracts with magnetic beads precoupled to Emi1 antibodies⁶ depleted more than 80% of Emi1 (Fig. 2b). Even though a small amount of Cdc20 immunoprecipitated with the anti-Emi1 beads, most Cdc20 (~80%) remained in the extracts (Fig. 2b, and data not shown). As shown in Fig. 2c–h, mock-depletion of CSF extracts did not affect cyclin B stability, metaphase arrest or sensitivity to calcium. In contrast, Emi1 depletion reproducibly induced cyclin B destruction and mitotic exit (Fig. 2c–h). Emi1-depleted extracts also degraded Mos in the absence of calcium (data not shown).

At time 0 in Emi1-depleted extracts, cyclin B levels were already lower than in mock-depleted extracts (Fig. 2c, f). Because cyclin B interacts with Emi1 (ref. 6), we needed to exclude the possibility that Emi1 depletion caused mitotic exit by depleting cyclin B as well. We found that only a small amount of cyclin B (~6.3%) was depleted together with Emi1 (Fig. 2b). Moreover, addition of the proteasome inhibitor MG-132 prevented the cyclin B decrease and mitotic exit in Emi1-depleted extracts (Supplementary Information, Fig. S2). Thus, the decrease in cyclin B at time 0 in Emi1-depleted extracts is most probably due to the occurrence of APC activation as Emi1 is depleted.

To verify that the cyclin B destruction and mitotic exit observed in Emi1-depleted extracts reflected the loss of Emi1's APC inhibitory function, we preincubated extracts with a stable carboxy-terminal fragment of Emi1 that is sufficient to inhibit APC activation *in vitro* and *in vivo*⁶. Pre-addition of this purified MBP–Emi1–CT protein rescued cyclin B stability and CSF metaphase arrest (Fig. 2f–h). Furthermore, the addition of purified Emi1 protein to CSF extracts

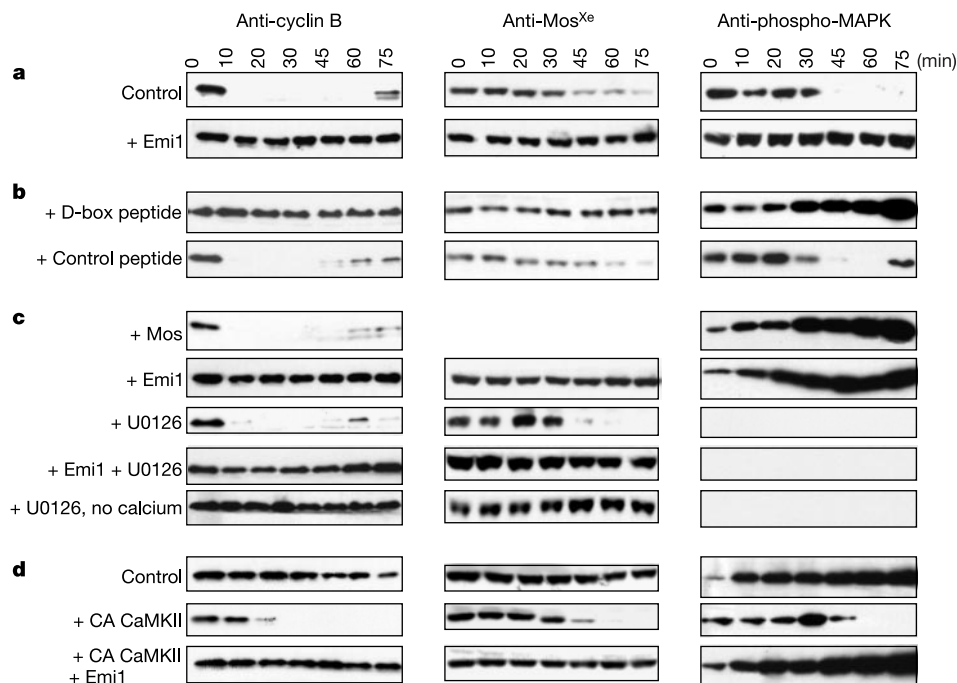


Figure 1 Emi1 is sufficient to prevent release from CSF block in the presence of Ca^{2+} . **a, b**, Addition of Emi1 or a D-box peptide stabilizes cyclin B and Mos, and blocks MAP kinase inactivation in the presence of calcium. CSF extracts were preincubated with MBP–Emi1 (1 μM) or MBP (1 μM , control), D-box peptide (675 μM) or control peptide (675 μM)⁶, and released with calcium. Samples were immunoblotted with antibodies against cyclin B2, Mos (Santa Cruz Biotechnology) and active MAP kinase (NEB). Times after calcium addition are indicated. **c**, Blocking by Emi1 of CSF release does not require

the MAP kinase pathway. CSF extracts were preincubated with MBP–Mos (1 μM), U0126 (50 μM) (Promega) or U0126 (50 μM) and then MBP–Emi1 (1 μM). Extracts were released with calcium (except where indicated) and analysed as in **a**. Control extracts in **a–c** re-entered mitosis by ~75 min. **d**, Constitutively active CaMKII cannot trigger calcium-independent cyclin B destruction and mitotic exit in the presence of Emi1. Extracts were analysed as in **a**. Times after CaMKII addition are indicated.

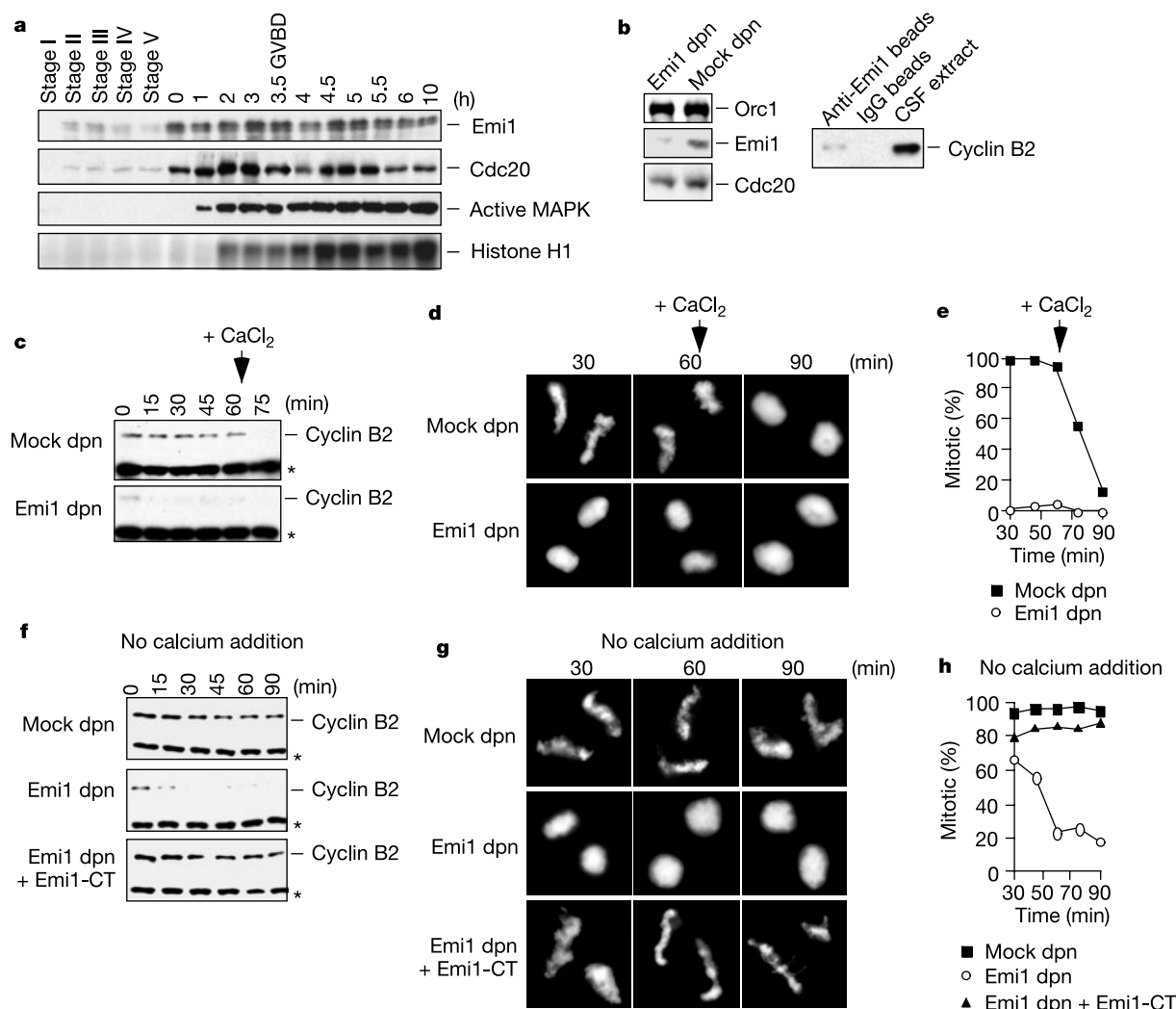


Figure 2 Emi1 is required for CSF arrest. **a**, Emi1 accumulates in the maturing oocyte. Samples from a maturation time course after progesterone addition (at time 0) and equal aliquots of stage I–V oocytes were analysed for Emi1, Cdc20 and active MAP kinase by immunoblotting and cyclin B-associated kinase activity by H1 kinase assay. GVBD, time at which 100% germinal vesicle breakdown had occurred. **b**, Specificity of Emi1 immunodepletion (dnp). Extracts were assayed by immunoblotting for Orc1 as a loading control, Emi1 and Cdc20. Beads from Emi1-depleted and mock-depleted extracts (from the equivalent of 1 μ l of extract) and CSF extract (1 μ l) were assayed for cyclin B by

immunoblotting (right). **c–h**, Emi1 depletion causes cyclin B destruction and mitotic exit in the absence of calcium. **c–e**, Emi1-depleted and mock-depleted extracts with **(d, e)** or minus **(c)** sperm DNA were warmed to 23 °C. Calcium was added at 60 min. Samples were taken at the indicated time after warming, then immunoblotted for cyclin B **(c)** or assayed for DNA morphology **(d)**. Mitotic figures were quantified **(e)**. **f–h**, Similar to **c–e**, except that one sample was preincubated with active non-degradable MBP–Emi1-CT domain during the immunodepletion, and no calcium was added. A background band (asterisks) served as a loading control; the exposure time was the same for all blots.

after partial depletion of Emi1 stabilized the remaining cyclin B and prevented mitotic exit (Supplementary Information, Fig. S3). Although it is possible that additional Emi1-interacting proteins important for CSF arrest are depleted together with Emi1, these data strongly indicate that Emi1 is required for the maintenance of CSF arrest in *Xenopus* eggs.

If Emi1 acts in CSF-arrested eggs to inhibit APC activation by Cdc20, then the addition of excess Cdc20 might be expected to overwhelm the inhibitory effect of endogenous Emi1 and thus cause APC activation and mitotic exit. To test this idea, we added increasing amounts of purified recombinant his-tagged Cdc20 to CSF-arrested extracts and examined whether these extracts exited from CSF arrest in the absence of calcium. Indeed, Cdc20 induced cyclin B degradation and mitotic exit in a dose-dependent fashion in CSF extracts. This effect was blocked by the addition of purified MBP–Emi1 protein but not by the addition of purified MBP–Mos protein (Fig. 3a–c). The added Mos protein did prolong mitosis as described previously²⁸.

Cyclin B protein is destroyed by ~10 min after egg activation²⁵ or release of CSF arrest in extracts (Fig. 1). Thus, the activation signal from CaMKII to the APC must be transduced quickly. Emi1 protein levels do not fluctuate significantly until the first mitosis after fertilization⁶, and Cdc20 protein is stable in the early embryo (Fig. 3d, e)⁴. How, therefore, is Emi1 prevented from inhibiting Cdc20 after fertilization? One possibility is that CaMKII activation leads to a change in binding between Emi1 and Cdc20. To test this idea, we analysed the binding of endogenous Emi1 and Cdc20 after calcium addition. Emi1 interacts specifically with Cdc20 *in vitro* and *in vivo*⁶. Before calcium addition (time 0), Cdc20 immunoprecipitated together with Emi1. By 2.5 min after calcium addition, only background amounts of Cdc20 binding were detectable (Fig. 3d). We saw a similar loss of binding when ³⁵S-labelled Cdc20 translated *in vitro* was added to CSF extracts and immunoprecipitated together with Emi1 (Fig. 3e).

The release of Cdc20 from Emi1 might result from a change in modification in either protein. After calcium addition we observed a

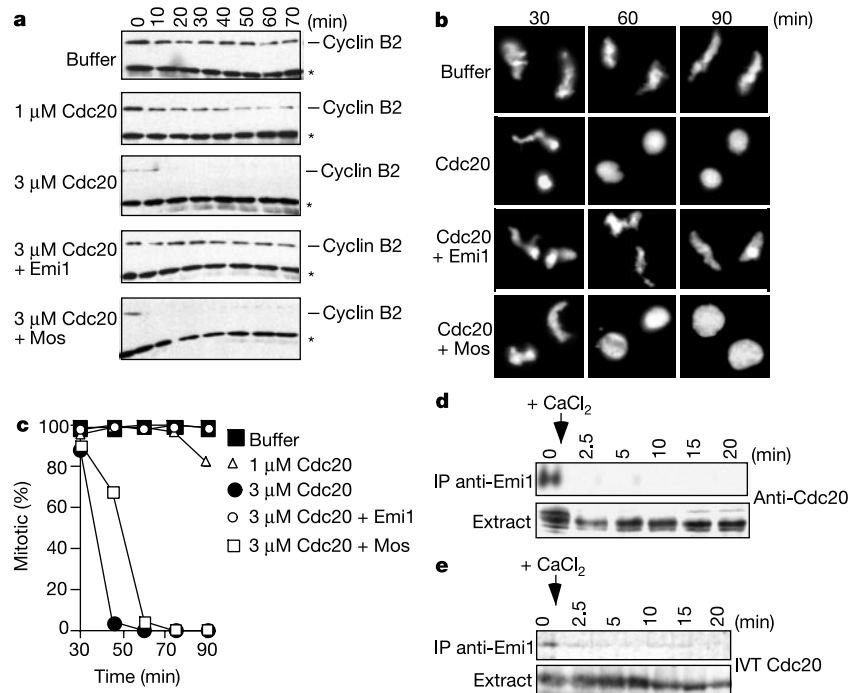


Figure 3 Addition of exogenous Cdc20 to CSF extracts activates cyclin B destruction and mitotic exit in the absence of calcium. **a**, CSF extracts were preincubated with buffer, 1 or 3 μ M Cdc20, 3 μ M Cdc20 preincubated with 3 μ M Emi1, or 3 μ M Cdc20 preincubated with 1 μ M Mos. Extracts were warmed to 23 °C and aliquots were analysed for cyclin B levels by immunoblotting. A background band (asterisks) served as a loading control; the exposure time was the same for all blots. **b**, **c**, Sperm DNA was added to extracts treated

as in **a**, and morphology was assessed (**b**) and quantified (**c**). **d**, **e**, Emi1 and Cdc20 dissociate after release from CSF arrest. Calcium was added at time 0 to CSF extracts (**d**) or CSF extracts preincubated with 35 S-labelled Cdc20 translated *in vitro* (IVT) (**e**), and the immunoprecipitation (IP) of Emi1 together with Cdc20 was assayed by immunoblotting or autoradiography. Unprecipitated extracts from specific time points were resolved by SDS-PAGE and analysed as loading controls (lower panels).

rapid increase in the electrophoretic mobility of Cdc20 that was consistent with dephosphorylation (Fig. 3d, e). Alternatively, Emi1 might be a target of CaMKII after its calcium-induced activation. Interestingly, Emi1 contains one consensus CaMKII phosphorylation site, as well as three additional less conserved phosphorylation sites.

Our data demonstrate that Emi1 is essential for the maintenance of CSF metaphase arrest. Masui originally proposed that CSF would (1) 'appear' (become activated) during oocyte maturation; (2) 'disappear' (become inactivated) after fertilization; (3) be inactivated under the same conditions as those that cause egg activation; (4) be sufficient for the metaphase of MII (CSF) arrest in eggs; and (5) cause a reversible mitotic block in the zygote¹. Unlike the Mos/MAPK pathway, which is not sufficient to maintain CSF arrest, Emi1 is both necessary and sufficient for metaphase of MII arrest. Thus, there seems to be a CSF biochemical pathway that acts to prevent cyclin B/Cdc2 inactivation until fertilization (Fig. 4). The Mos/MAPK pathway acts to keep Cdc2 active through positive regulation and by increasing cyclin B synthesis during the MI–MII transition (reviewed in ref. 29), whereas Emi1 acts to prevent cyclin B destruction through APC inhibition in MII. It remains to be seen how Emi1 becomes activated during oocyte maturation, whether Emi1 has a role earlier in oocyte maturation, and whether the Mos/MAPK pathway regulates Emi1 during this time. □

Methods

Plasmids and proteins

Emi1 and Mos constructs were as described^{6,8} and proteins were produced and purified with the use of standard protocols. To prepare constitutively active CaMKII, residues 1–290 were subcloned from rat brain CaMKII α (H. Schulman) into pCS2 vector.

Xenopus extracts and oocytes

Xenopus CSF extracts were prepared as described³⁰ and 400 μ M CaCl_2 was used to release extracts from the CSF state. For addition experiments, CSF extracts were preincubated

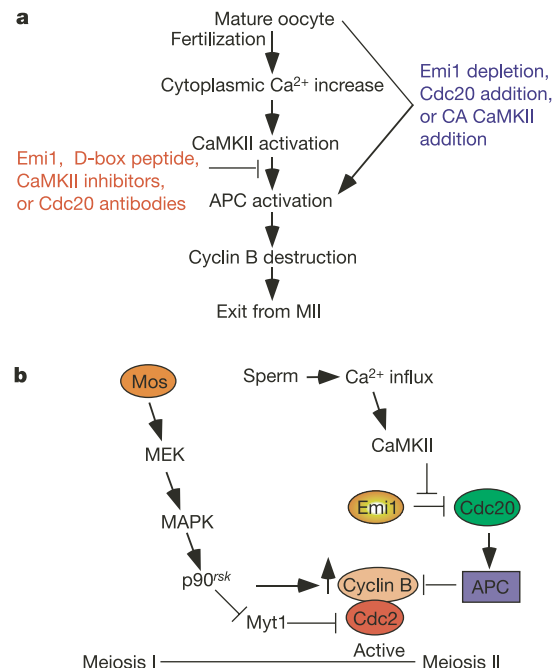


Figure 4 Model. **a**, Perturbations that bypass the calcium requirement for APC activation in the mature CSF-arrested oocyte: Emi1 depletion, excess Cdc20 protein, or the addition of constitutively active CaMKII (ref. 26). Calcium-triggered APC activation is blocked by the addition of Emi1 protein, Cdc20-neutralizing antibodies⁴ or CaMKII inhibitors²⁶. **b**, Model of how Emi1 and the Mos/MAPK pathway act to prevent release from metaphase of MII by stabilizing cyclin B/Cdc2 activity. Mos promotes cyclin B/Cdc2 activity by negatively regulating the Myt1 kinase and by increasing cyclin B synthesis during the MI–MII transition^{27,29}. Emi1 inhibits APC-dependent cyclin B destruction in MII, thereby stabilizing cyclin B/Cdc2 activity.

(15 min, 4°C) with the indicated reagent. Extracts (with or without calcium addition at time 0) were warmed to 23°C, and equal aliquots were taken at the indicated times.

For experiments using the addition of constitutively active CaMKII, CSF extracts were preincubated (15 min, 4°C) with buffer or MBP–Emi1 (300 nM). Constitutively active CaMKII translated *in vitro* was added (1:14) at time 0, extracts were warmed to 23°C without the addition of calcium, and equal aliquots were taken at the indicated times.

For oocyte maturation experiments, stage VI oocytes were treated with 5 µg ml⁻¹ progesterone and equal aliquots were collected and lysed over 10 h as described³¹. The cyclin-B-associated H1 kinase assay was as described³⁰.

Depletion experiments

Affinity-purified anti-Emi1 rabbit antibodies (5 µM) or anti-Emi1 mouse polyclonal serum (0.2 µl serum per µl extract) were bound to magnetic protein A–dynabeads or protein G–dynabeads, respectively (Dyna). Beads were washed (20 mM HEPES, pH 7.7, 100 mM KCl) and incubated with CSF extracts (1 h, 4°C), and samples were cleared of beads with a magnet. The process was repeated twice more (30 min, 4°C) and the triple-depleted extracts were warmed to 23°C for analysis. Mock-depletions were performed in exactly the same manner, except that purified rabbit IgG (5 µM) or mouse preimmune serum (0.2 µl serum per µl extract) was used instead. For DNA morphology analysis, sperm DNA was added (1,000 µl⁻¹) and DNA was analysed by staining with Hoechst 33258 (ref. 6). MBP–Emi1–CT domain (1 µM) was added for rescue experiments. We performed these Emi1 depletion experiments 10 different times with anti-Emi1–depleting antibodies (1) from several different rabbits; (2) against full-length Emi1 and an amino-terminal Emi1 fragment; (3) with affinity-purified versus crude sera; and (4) with anti-Emi1 sera from several different mice.

Immunoprecipitation assays

Calcium (400 µM) was added at time 0 to CSF extracts or CSF extracts preincubated (15 min, 4°C) with ³⁵S-labelled Cdc20 translated *in vitro*. Extracts were warmed to 23°C and samples were taken at the indicated times after calcium addition; they were then flash-frozen. Samples were diluted 1:10 in RIPP buffer⁶, precleared with protein G–Sepharose, incubated with anti-Emi1 serum (1 h, 4°C), bound to protein G–Sepharose (45 min, 4°C) and washed four times in NETN buffer (20 mM Tris–HCl pH 7.5, 150 mM NaCl, 0.5% Nonidet P40, 1 mM dithiothreitol, 1 mM EDTA, 1% aprotinin); bound Cdc20 was resolved by SDS–polyacrylamide gel electrophoresis and immunoblotting or autoradiography.

Received 10 October 2001; accepted 21 February 2002.

- Masui, Y. The elusive cytotstatic factor in the animal egg. *Nature Rev. Mol. Cell Biol.* **1**, 228–232 (2000).
- Masui, Y. & Markert, C. L. Cytoplasmic control of nuclear behaviour during meiotic maturation of frog oocytes. *J. Exp. Zool.* **177**, 129–145 (1971).
- Murray, A. W., Solomon, M. J. & Kirschner, M. W. The role of cyclin synthesis and degradation in the control of maturation promoting factor activity. *Nature* **339**, 280–286 (1989).
- Lorca, T. *et al.* Fizzy is required for activation of the APC/cyclosome in *Xenopus* egg extracts. *EMBO J.* **17**, 3565–3575 (1998).
- Vorlauffer, E. & Peters, J. M. Regulation of the cyclin B degradation system by an inhibitor of mitotic proteolysis. *Mol. Biol. Cell* **9**, 1817–1831 (1998).
- Reimann, J. D. *et al.* Emi1 is a mitotic regulator that interacts with Cdc20 and inhibits the anaphase promoting complex. *Cell* **105**, 645–655 (2001).
- Sagata, N., Daar, I., Oskarsson, M., Showalter, S. D. & Vande Woude, G. F. The product of the *mos* proto-oncogene as a candidate 'initiator' for oocyte maturation. *Science* **245**, 643–646 (1989).
- Yew, N., Mellini, M. L. & Vande Woude, G. F. Meiotic initiation by the *mos* protein in *Xenopus*. *Nature* **355**, 649–652 (1992).
- Colledge, W. H., Carlton, M. B., Udy, G. B. & Evans, M. J. Disruption of *c-mos* causes parthenogenetic development of unfertilized mouse eggs. *Nature* **370**, 65–68 (1994).
- Hashimoto, N. *et al.* Parthenogenetic activation of oocytes in *c-mos*-deficient mice. *Nature* **370**, 68–71 (1994).
- Furuno, N. *et al.* Suppression of DNA replication via *Mos* function during meiotic divisions in *Xenopus* oocytes. *EMBO J.* **13**, 2399–2410 (1994).
- Gross, S. D. *et al.* The critical role of the MAP kinase pathway in meiosis II in *Xenopus* oocytes is mediated by p90^{Rsk}. *Curr. Biol.* **10**, 430–438 (2000).
- Bhatt, R. R. & Ferrell, J. E. Jr protein kinase p90 Rsk as an essential mediator of cytotstatic factor activity. *Science* **286**, 1362–1365 (1999).
- Gross, S. D., Schwab, M. S., Lewellyn, A. L. & Maller, J. L. Induction of metaphase arrest in cleaving *Xenopus* embryos by the protein kinase p90^{Rsk}. *Science* **286**, 1365–1367 (1999).
- Haccard, O. *et al.* Induction of metaphase arrest in cleaving *Xenopus* embryos by MAP kinase. *Science* **262**, 1262–1265 (1993).
- Shibuya, E. K. & Ruderman, J. V. *Mos* induces the *in vitro* activation of mitogen-activated protein kinases in lysates of frog oocytes and mammalian somatic cells. *Mol. Biol. Cell* **4**, 781–790 (1993).
- King, R. W. *et al.* A 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. *Cell* **81**, 279–288 (1995).
- Sudakin, V. *et al.* The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. *Mol. Biol. Cell* **6**, 185–197 (1995).
- Shah, J. V. & Cleveland, D. W. Waiting for anaphase: Mad2 and the spindle assembly checkpoint. *Cell* **103**, 997–1000 (2000).
- Chen, R. H., Waters, J. C., Salmon, E. D. & Murray, A. W. Association of spindle assembly checkpoint component XMAP2 with unattached kinetochores. *Science* **274**, 242–246 (1996).
- Sharp-Baker, H. & Chen, R. H. Spindle checkpoint protein Bub1 is required for kinetochore localization of Mad1, Mad2, Bub3, and CENP-E, independently of its kinase activity. *J. Cell Biol.* **153**, 1239–1250 (2001).
- Abrieu, A. *et al.* Mps1 is a kinetochore-associated kinase essential for the vertebrate mitotic checkpoint. *Cell* **106**, 83–93 (2001).
- Sagata, N., Watanabe, N., Vande Woude, G. F. & Ikawa, Y. The *c-mos* proto-oncogene product is a

- cytotstatic factor responsible for meiotic arrest in vertebrate eggs. *Nature* **342**, 512–518 (1989).
- Castro, A. *et al.* Cyclin B/cdc2 induces *c-Mos* stability by direct phosphorylation in *Xenopus* oocytes. *Mol. Biol. Cell* **12**, 2660–2671 (2001).
- Watanabe, N., Hunt, T., Ikawa, Y. & Sagata, N. Independent inactivation of MPF and cytotstatic factor (*Mos*) upon fertilization of *Xenopus* eggs. *Nature* **352**, 247–248 (1991).
- Lorca, T. *et al.* Calmodulin-dependent protein kinase II mediates inactivation of MPF and CSF upon fertilization of *Xenopus* eggs. *Nature* **366**, 270–273 (1993).
- Taieb, F. E., Gross, S. D., Lewellyn, A. L. & Maller, J. L. Activation of the anaphase-promoting complex and degradation of cyclin B is not required for progression from meiosis I to II in *Xenopus* oocytes. *Curr. Biol.* **11**, 508–513 (2001).
- Guadagno, T. M. & Ferrell, J. E. Jr Requirement for MAPK activation for normal mitotic progression in *Xenopus* egg extracts. *Science* **282**, 1312–1315 (1998).
- Ferrell, J. E. Jr *Xenopus* oocyte maturation: new lessons from a good egg. *BioEssays* **21**, 833–842 (1999).
- Murray, A. W. Cell cycle extracts. *Methods Cell Biol.* **36**, 581–605 (1991).
- Sahaskey, M. L. & Ferrell, J. E. Jr Distinct, constitutively active MAPK phosphatases function in *Xenopus* oocytes: implications for p42 MAPK regulation *in vivo*. *Mol. Biol. Cell* **10**, 3729–3743 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank T. Hunt for Xcyclin B antibodies; E. Kramer and J. Peters for XCdc20 antibodies and construct; H. Schulman for CaMKIIα cDNA; H. Yu for D-box peptide; Ferrell laboratory members for technical assistance; and J. Peters and J. Ferrell for comments on the manuscript.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.J. (e-mail: pjackson@stanford.edu).

Dominant role of the niche in melanocyte stem-cell fate determination

Emi K. Nishimura*†‡, Siobhán A. Jordan§, Hideo Oshima‡||, Hisahiro Yoshida*, Masatake Osawa¶, Mariko Moriyama¶, Ian J. Jackson§, Yann Barrandon||, Yoshiki Miyachi† & Shin-Ichi Nishikawa*¶

* Department of Molecular Genetics and † Department of Dermatology, Graduate School of Medicine, Kyoto University, Shogoin Kawaharacho 53, Sakyo-ku, Kyoto 606-8507, Japan
§ MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK
|| Ecole Normale Supérieure, 46 rue d'Ulm, Paris 75230, Cedex 05, France
¶ Riken Center for Developmental Biology, Minatogima-minamicho 2-2-3, Chuo-ku, Kobe, 650-0047, Japan

Stem cells—which have the capacity to self-renew and generate differentiated progeny—are thought to be maintained in a specific environment known as a niche^{1–3}. The localization of the niche, however, remains largely obscure for most stem-cell systems. Melanocytes (pigment cells) in hair follicles proliferate and differentiate closely coupled to the hair regeneration cycle⁴. Here we report that stem cells of the melanocyte lineage can be identified, using *Dct-lacZ* transgenic mice^{5,6}, in the lower permanent portion of mouse hair follicles throughout the hair cycle. It is only the population in this region that fulfils the criteria for stem cells, being immature, slow cycling, self-maintaining and

‡ Present addresses: Division of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Dana 630A, 44 Binney Street, Boston, Massachusetts 02115, USA (E.K.N.); Department of Plastic & Reconstructive Surgery, St Marianna University School of Medicine, 2-16-1 Sugao Miyamae-ku, Kawasaki, 216-8511, Japan (H.O.).

fully competent in regenerating progeny on activation at early anagen (the growing phase of hair follicles). Induction of the repigmentation process in *K14-steel factor* transgenic mice⁷ demonstrates that a portion of amplifying stem-cell progeny can migrate out from the niche and retain sufficient self-renewing capability to function as stem cells after repopulation into vacant niches. Our data indicate that the niche has a dominant role in the fate determination of melanocyte stem-cell progeny.

Hair regeneration initiates at early anagen from the bulge area, where stem cells of the follicular keratinocytes reside^{8–10} (Fig. 1a). This is followed by downward growth of the basal portion of the hair follicles, the hair matrix. Subsequently, these follicles regress at

catagen and become resting at telogen¹¹. Melanocytes in the hair matrix proliferate in anagen, differentiate to produce melanin pigment that is transferred to hairs, and then die by apoptosis during catagen¹². The stem-cell system is divided into three compartments: stem cells, transiently amplifying cells and mature cells¹³. The latter two compartments of the melanocyte lineage reside in the hair matrix. Although various models have been proposed for the localization of the melanocyte stem-cell compartment^{14–18}, no previous studies have succeeded in identifying a melanoblast population that fulfils the criteria for stem cells¹³.

We used two tools to determine the localization of melanocyte stem cells. The first tool is a transgenic mouse carrying the *lacZ* reporter under the control of the *Dct* promoter. *Dct* codes for the enzyme dopachrome tautomerase (also known as TRP-2), an early marker of the melanocyte lineage. *Dct-lacZ* expression in the transgenic mice is colocalized with the *Dct* messenger RNA expression (data not shown), and this allowed us to detect *Dct*⁺ cells with high sensitivity. Secondly, we used an antagonistic monoclonal antibody, ACK2, to block the function of Kit (another marker of the melanocyte lineage)¹⁹. We have demonstrated previously that treatment of mice with ACK2 selectively depletes amplifying populations of melanoblasts, spermatocytes and haematopoietic cells^{19–22}, leaving resting cells intact. Melanoblasts colonize developing hair follicles perinatally after migration through the dermis and epidermis^{23,24}. In this study, when neonatal mice were treated with ACK2, the first hairs to develop were almost all unpigmented (Fig. 1b, c). However, in the subsequent hair cycle, hair pigmentation was restored in dorsal over hairs²⁵ (20% of dorsal hair) and ventral sensory hairs (2% of ventral hair) (Fig. 1d), which develop earlier than other hairs²⁶ (Fig. 1e), indicating the existence of Kit-independent precursors somewhere in these follicles. Indeed, *Dct-lacZ*⁺ cells are found exclusively in the outer root sheath (ORS) of the lower permanent portion: the bulge area where arrector pili muscle is attached and occasionally also immediately below the area (sub-bulge area) of these follicles in ACK2-treated animals (Fig. 1g). In control animals, *Dct-lacZ*⁺ cells are also abundant in the hair matrix (Fig. 1f). During anagen of the next hair cycle, pigmented melanocytes were found in the hair matrix of dorsal over hairs and ventral sensory hairs, whose bulge area retains *lacZ*⁺ cells (Fig. 1h).

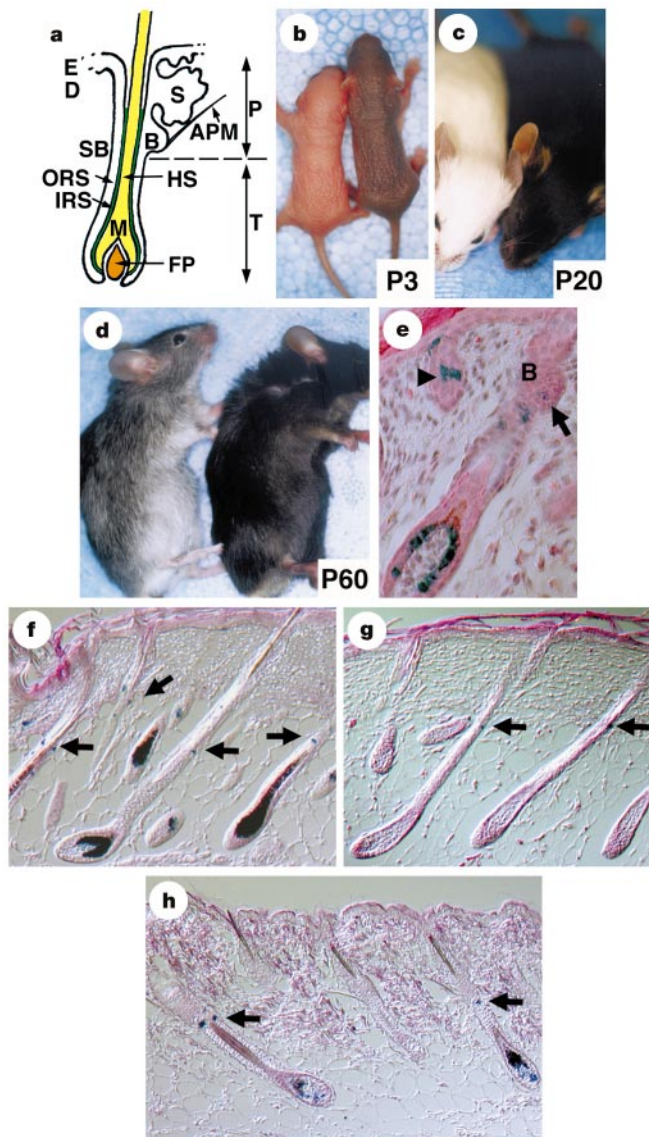


Figure 1 *Dct-lacZ*⁺ cells in the bulge area survive in a Kit-independent manner. **a**, The hair follicle structure. B, bulge region; SB, sub-bulge region; ORS, outer root sheath; HS, hair shaft; E, epidermis; D, dermis; IRS, inner root sheath; S, sebaceous gland; APM, arrector pili muscle; M, hair matrix; FP, follicular papilla; P, permanent portion; T, transient portion. **b–d**, Coat colour of neonatally ACK2-treated (left) and -untreated (right) mice. **e–h**, Localization of *lacZ*⁺ melanoblasts (arrows) in hair follicles of control mice at P0.5 (**e**) and P3 (**f**), and of ACK2-treated mice at P3 (**g**) and P25 (**h**). The arrowhead in **e** indicates a melanoblast in a late-developing hair bud. Note that *LacZ*⁺ cells exist only in the bulge area of ACK2-treated over-hair follicles (**g**), and only these follicles recover hair matrix melanocytes in the next hair cycle (**h**). Original magnification: $\times 100$ (**e**); $\times 200$ (**f–h**).

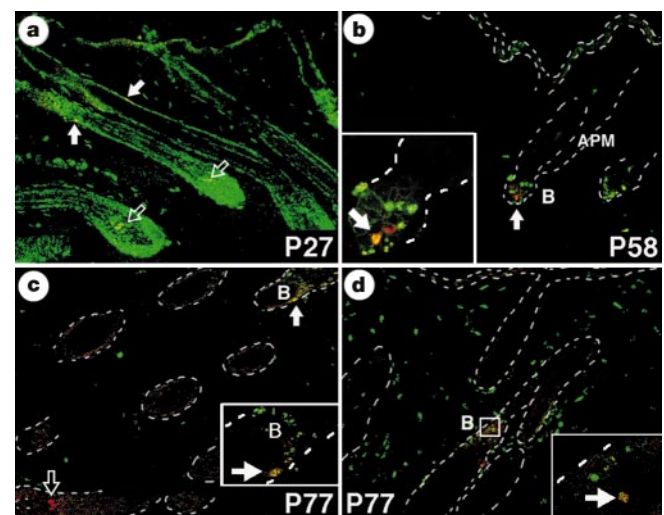


Figure 2 Bulge melanoblasts retain BrdU during hair cycling. *Dct-lacZ* transgenic mice were injected subcutaneously with BrdU ($10 \mu\text{g g}^{-1}$ body weight) twice daily from P20 to P27 (**a–c**) or from P0 to P7 (**d**). Red, β -galactosidase; green, BrdU. **a**, Almost all cells were labelled with BrdU at P27 just after injection. **b–d**, At P58 and P77, BrdU label was retained only in the cells in the bulge area: keratinocytes and *lacZ*⁺ melanoblasts (filled arrow), but not in hair matrix melanocytes (open arrow). B, bulge area; APM, arrector pili muscle. Original magnification, $\times 200$ (inset $\times 500$).

Table 1 Pigmented hair formation from implants of late-catagen vibrissal follicles

Origin of implants	Hair regeneration	Pigmented hair formation
Upper permanent portion	6/8	0/6
Lower permanent portion	5/8	3/5
Hair bulb	3/6	0/3

Furthermore, $lacZ^{+}$ cells colonized the bulge area of these early developed follicles by postnatal day (P) 0.5 (Fig. 1e), and survived in a Kit-independent manner with low or undetectable levels of Kit expression. These melanoblasts in the bulge-sub-bulge area may represent the stem-cell compartment of the melanocyte lineage. To be positively identified as stem cells, these cells need to show attributes common to the stem-cell compartment. Thus, we examined their rate of proliferation and their functional competence to regenerate the melanocyte lineage in new hair follicles.

To evaluate the proliferation rate, cells were labelled with 5-bromo-2-deoxyuridine (BrdU) during early-mid anagen in *Dct-lacZ* transgenic mice. Using this protocol, most epidermal and follicular cells including $lacZ^{+}$ cells (arrows in Fig. 2) are labelled by BrdU (Fig. 2a). The label was chased for 26–70 days (over one or two hair cycles), followed by immunofluorescent staining for BrdU. At P58, only cells in the bulge region of telogen hair follicles, including $lacZ^{+}$ melanoblasts, retained the label in their nuclei (Fig. 2b). Even after 70 days of chase, $lacZ^{+}$ bulge melanoblasts retained BrdU (Fig. 2d), whereas hair matrix melanocytes have lost the BrdU label (Fig. 2c). Furthermore, $lacZ^{+}$ melanoblasts in the bulge could be labelled only at the beginning of early anagen. These findings indicate that bulge melanoblasts are slow-cycling cells activated only at early anagen, thereby possessing an attribute of stem cells.

To determine whether the melanoblasts in the lower permanent portion are competent to regenerate melanocytes in the hair follicle, we used the hair reconstitution technique with vibrissal (whisker)

hair follicles⁹. *Dct-lacZ*⁺ melanoblasts were found preferentially in the ORS of the lower permanent portion of vibrissal follicles throughout hair cycling (Fig. 3a; and data not shown). By contrast, mature melanocytes were found in the hair matrix at anagen. We divided the follicles into three fragments as shown in Fig. 3a. We implanted individual fragments of follicles from ROSA 26 mice²⁷, which express a *lacZ* reporter gene in most tissues including the follicular epithelium⁹, into the skin of albino neonates. The regeneration of chimaeric $lacZ^{+}$ follicles was obtained when any of the three fragments of late catagen or the upper permanent portion of follicles of any hair-cycle stage was implanted (Table 1 and ref. 9). However, only implants from the lower permanent portion gave pigmented hair (Fig. 3b–d). In contrast, neither the upper permanent portion nor the hair bulb of any stage developed pigmented hairs ($N = 25$ and 17, respectively). These findings demonstrate that melanocyte stem cells are located only in the lower permanent portion of vibrissal follicles.

We next analysed the behaviour of melanocyte lineage cells in pelage (body) hair cycling chronologically. In *Dct-lacZ* transgenic mice, at least one $lacZ^{+}$ cell is maintained in the lowest permanent portion throughout a hair's cycle (Fig. 4a–i). The population usually consists of small oval-shaped cells containing neither melanin granules nor tyrosinase, indicating that bulge melanoblasts are immature (Fig. 4j–l; and data not shown). At the transition from telogen to anagen, the *lacZ* expression level and cell size of melanocyte lineage cells increase (Fig. 4a, b), and they eventually divide (Fig. 4c). At least one $lacZ^{+}$ cell remains in the bulge (Fig. 4d–g). Others, surrounded by the growing hair germ, extend their processes towards the follicular papillae (Fig. 4d–f, m). These two populations are segregated away from each other in the growing follicle (Fig. 4f, g). The latter population localizes to the hair matrix, undergoes cell divisions, and differentiates into pigmented melanocytes (Fig. 4f, g, h).

The cell behaviour described above, taken together with BrdU retention and reconstitution experiments, demonstrates that melanoblasts in the lower permanent portion are undifferentiated slow-cycling cells, that they self-maintain, and that they supply large

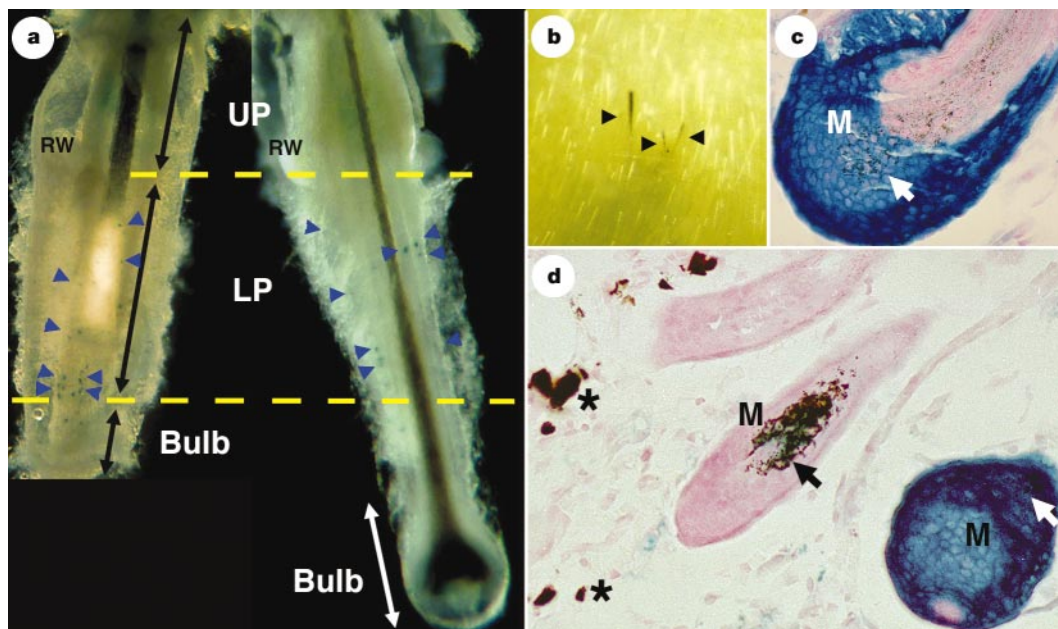


Figure 3 Regeneration of functional melanocytes from an implanted lower permanent portion. **a**, Distribution of $lacZ^{+}$ cells in vibrissal follicles of *Dct-lacZ* transgenic mice. Left, late catagen (stage 5); right, anagen. $lacZ^{+}$ melanin[−] cells are located exclusively in the lower permanent portion (LP) (blue arrowheads). UP, upper permanent portion; RW, ringwulst. **b–d**, Regenerated chimaeric, pigmented hairs: the lower permanent portion of

a ROSA 26-derived vibrissal follicle was implanted into the albino skin (OF1) together with charcoal. **c**, **d**, β -galactosidase staining of the P35 skin shows that the lower permanent portion develops pigmented $lacZ^{+}$ chimaeric follicles (arrows). M, hair matrix. Charcoal is indicated by asterisks. Original magnification: $\times 630$ (**c**); $\times 200$ (**d**).

numbers of differentiated progeny to the hair matrix. These features satisfy the criteria for a stem cell¹³. Whereas only one to a few lacZ⁺ cells were detected in the bulge area of under-hair follicles, more were found in the bulge–sub-bulge area of over-hair follicles. This

lower permanent portion, which serves as the niche for melanocyte stem cells, appears to vary in size among hair types but always lies on or immediately below the niche for follicular keratinocytes.

The above observations suggest the involvement of extrinsic

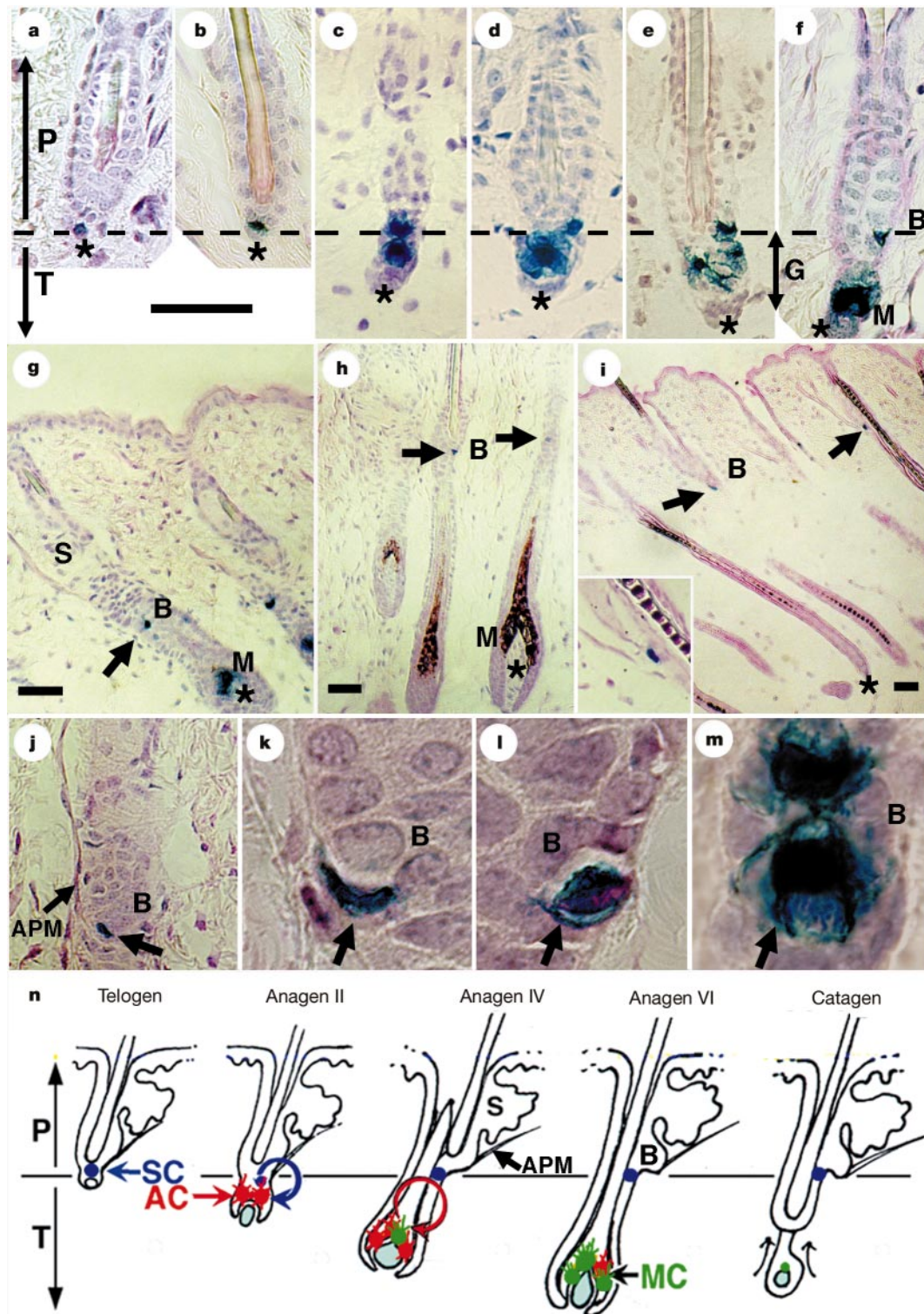


Figure 4 LacZ⁺ melanoblast (arrows) distribution during hair cycling. **a**, Telogen; **b**, anagen I; **c**, anagen II; **d**, late anagen II; **e**, anagen III; **f**, late anagen III; **g**, anagen IV; **h**, anagen V–VI; and **i**, catagen. The dotted line indicates the bulge area, and asterisks indicate follicular papilla. P, permanent portion; T, transient portion; G, hair germ; B, bulge area; M, hair matrix. Scale bars, 50 μ m. **j–m**, Resting melanoblasts in the bulge of anagen V–VI follicles (**j–l**) are small and round compared with activated melanoblasts in

the bulge of anagen II (**m**). S, sebaceous gland; APM, arrector pili muscle. **n**, Summary of melanocyte behaviour during hair cycling. Stem cells (blue), maintained in the lowest permanent portion throughout the hair cycle, are reactivated at early anagen to supply amplifying progeny (red) to the hair matrix, where most of them mature into differentiated melanocytes (green). SC, stem cells; AC, amplifying cells; MC, mature cells. Original magnification: $\times 400$ (**j**); $\times 1,000$ (**k–m**).

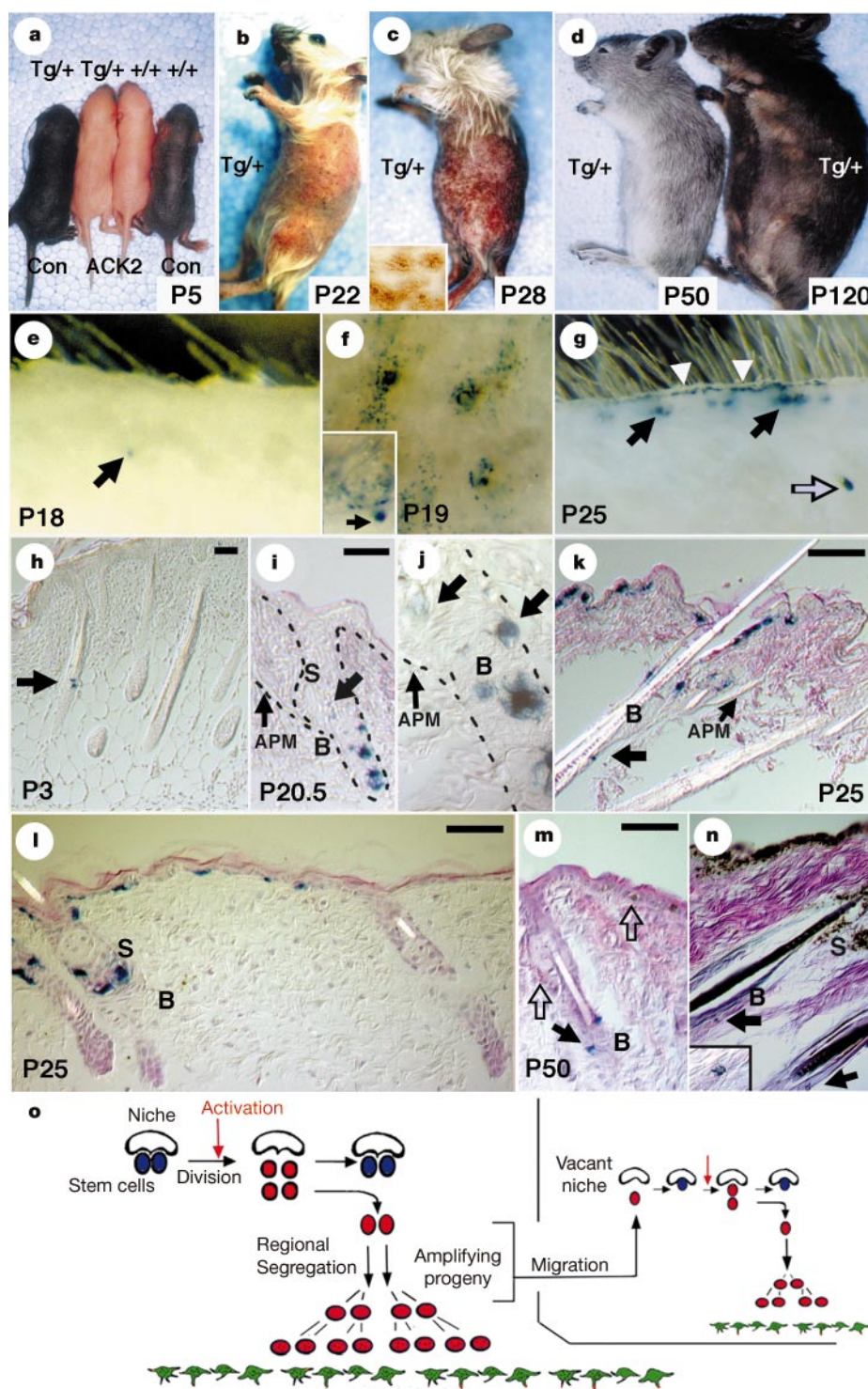


Figure 5 Stem cells recolonize empty niches. *K14-SLF, Dct-LacZ* double transgenic (*Tg/+*) (**a–n**) and *Dct-LacZ* transgenic mice (abbreviated as *+/+* in **a**) were treated with ACK2 at neonatal stage. **a–d**, Pigmentation pattern of ACK2-treated mice. The recovered coat colour pattern after first moulting was identical between *Tg/+* and *+/+* mice (**d**; see also Fig. 1d, left). After second or third moulting, all hairs eventually become pigmented in the *Tg/+* skin (**d**, right), but not in *+/+* skin (see also Fig. 1d, left; and data not shown). **e–n**, β -galactosidase staining of ACK2-treated *Tg/+* skin. **e–g**, Whole-mount views show the radial pattern of melanoblast migration from the *lacZ*⁺ cell-retaining follicles: ventral sensory-hair follicles (**e**, **f**) and dorsal over-hair follicles (**g**). Views of cross-sections (**e**, **f** inset, **g**) and skin surface (**f**). Filled and open arrows indicate melanocytes in the permanent portion and hair matrix of the follicle, respectively. Arrowheads (**g**) indicate melanocyte spreading in the epidermis. **h–n**, Distribution of spreading *lacZ*⁺ cells to

adjacent follicles was examined histologically. In the first hair cycling, *lacZ*⁺ cells are present in the bulge area of dorsal over-hair follicles (**h**). During the second hair cycling, *lacZ*⁺ cells are found in the upper permanent portion and hair matrix of the follicles (**i**, **j**) and subsequently also in the surrounding epidermis (**k**). Borders between colonized and non-colonized skins are found (**l**). Whereas just after colonization melanoblasts in the epidermis and upper permanent portion have not yet pigmented (**l**), they eventually become pigmented (**m**, open arrows) and supply melanin granule to the surrounding keratinocytes (**n**; P120). Note that *lacZ*⁺ melanin[−] cells (filled arrows and inset), but no pigmented cells, are present in the bulge. S, sebaceous gland; APM, arrector pili muscle; B, bulge area; D, dorsal skin; V, ventral skin. Scale bars, 50 μ m. **o**, Niche-based stem-cell model. Blue, resting stem cells; red, amplifying progeny; green, mature cells.

factors in segregation of the stem-cell compartment from other compartments. Indeed, progeny of melanocyte stem cells proliferate in the hair matrix located close to the follicular papillae in which *steel factor* (*SLF*) (the ligand of Kit receptor) is expressed²⁸. Transgenic mice in which *SLF* is constitutively expressed from the keratin (*K14*) promoter by basal keratinocytes have pigmented skin resembling human skin, which also expresses *SLF* in the epidermis. Using *K14-SLF* transgenic mice (*K14-SLF*/+), we have demonstrated previously that *SLF* expression alone is sufficient to provide the interfollicular epidermis with an ability to support the presence of melanocytes²⁹.

To determine where the reservoir of epidermal melanocytes exists, we depleted the Kit-dependent compartment from *K14-SLF*/+; *Dct-lacZ*/+ transgenic mouse (referred to hereafter as Tg/+) skin by neonatal ACK2 treatment, and analysed the subsequent pattern of re-pigmentation. Treated animals showed white hairs (Fig. 5a–c) and white skin (Fig. 5a). LacZ⁺ cells were maintained exclusively in the bulge area of early developed follicles: dorsal over-hair follicles (Fig. 5h) and ventral sensory-hair follicles (Fig. 5e) during the first hair cycle, identical to the pattern of ACK2-treated +/+; *Dct-lacZ*/+ animals (Fig. 1g). Thus, ACK2 treatment resulted in a large number of vacant niches both in Tg/+ and in +/+ animals. The absence of functional melanoblasts in these follicles was confirmed by the finding that almost all hairs grown in these types of follicles are unpigmented in the next hair cycle (Fig. 5d left, m). On the other hand, the early developed follicles that maintain melanoblasts in the bulge grow pigmented hairs. Interestingly, in Tg/+ mice, epidermal re-pigmentation appeared to start from such hair follicles (Fig. 5b, c, f).

We analysed the re-pigmentation process in Tg/+ mice with β -galactosidase staining of whole-mounted skins. In the ventral skin, lacZ⁺ melanoblasts were detected in the permanent portion of sensory-hair follicles (indicated by an arrow) at P18 and P19 (Fig. 5e, f (inset)), and some were already in their surrounding epidermis (Fig. 5f). Indeed, histological examination of the Tg/+ skin demonstrated that the progeny of surviving melanoblasts in the bulge can migrate upwards to the epidermis in the presence of SLF (Fig. 5i–k). These findings indicate that bulge stem cells are the source of melanocytes in the epidermis. Increasing numbers of lacZ⁺ cells appeared in the epidermis in a concentric pattern around the follicles (Fig. 5f). These cells were undergoing differentiation, leading to pigmented spots on the skin (Fig. 5b), which with time enlarged and fused (Fig. 5c). These observations suggest that *SLF* expression in the epidermis provides new routes connecting the follicular ORS and the epidermis, along which melanoblasts can migrate (Fig. 5f, g). Taking this route, lacZ⁺ cells eventually colonized almost all of the empty bulge area of follicles over the whole skin in a few hair cycles (Fig. 5d, m, n, and data not shown).

The whole process indicates that a population of melanocyte stem-cell progeny, which migrates out from the niche, amplifies outside the niche and colonizes vacant niches, eventually functioning as stem cells. After repopulation into niches they return to a quiescent state with small and round cell bodies (Fig. 5m, n), and are reactivated at the next early anagen (data not shown). Amplified melanoblasts that settle outside the niche differentiate into pigmented melanocytes (Fig. 5m, n). Eventually almost all white hairs are replaced with new, pigmented hairs (Fig. 5d, right). Interestingly, the coat colour pattern restored in the second hair cycle remained unchanged in the absence of *K14-SLF* transgene (Fig. 1d, left; and data not shown). Given that restoration (Fig. 5) started from 2% of the follicles in the Tg/+ ventral skin, amplified progeny from these follicles formed 50 times more new stem-cell systems in surrounding vacant follicles, indicating their extremely high self-maintaining ability. In the compartment model⁸, the amplifying progeny, which migrated out from the niche, would correspond to the 'transiently amplifying compartment'. Our data indicate that some melanocytes in this compartment can return to the stem-cell compartment when

they populate the 'niche'.

We have identified melanocyte stem cells in the lower permanent portion of hair follicles, and have demonstrated the potent capacity of this area to serve as a niche wherein amplifying stem-cell progeny assumes a stem-cell fate (Fig. 5o). The re-pigmentation process we observed strongly suggests that human melanoblasts in the ORS are the stem-cell source for hair matrix melanocytes and for epidermal melanocytes. Indeed, the pigmentation pattern is reminiscent of the recovery process of human vitiligo, an acquired disease characterized by loss of melanocytes³⁰. □

Methods

Animals and antibody treatment

We obtained ROSA 26 mice (C57BL/6) from the Jackson laboratory. OF1 (albino) and athymic (nu/nu) mice were from Iffa Credo. *Dct-lacZ* transgenic mice, *K14-SLF* transgenic mice and ACK2 have been described previously^{5,7,19}. Neonatal mice were treated subcutaneously with 0.2 mg ACK2 at P0, P2 and P4.

β -galactosidase staining

Skin samples from *Dct-lacZ* transgenic mice were immersed in fixation solution (2% formaldehyde, 0.2% glutaraldehyde, 0.02% Nonidet P40 in phosphate-buffered saline (PBS) at pH 7.4), irradiated for 20 s in a 600 W microwave oven and kept on ice for 60 min. The fixed samples were stained in halogenated indolyl- β -D-galactoside (Bluo-Gal; Gibco-BRL) solution. The snap-frozen sections were post-fixed and counterstained with eosin and observed by differential interference contrast.

Immunohistochemistry

Skin samples were immersed in 2% paraformaldehyde/PBS (pH 7.4) and irradiated in a 600 W microwave oven for 60 s at 4 °C. The fixed skin was embedded in OCT compound and snap frozen. For double immunofluorescent staining, 14 mm cryosections were processed in the following solutions: PBSMT (2% skim milk powder and 0.1% v/v Triton X-100 in PBS) for 20 min at room temperature; 10 mg ml⁻¹ rat anti-Kit (ACK2) (1:100) or fluorescein isothiocyanate (FITC)-conjugated mouse anti-BrdU (Becton Dickinson) (1:5), and rabbit anti- β -galactosidase (Cappel) (1:900) in PBST (0.1% v/v Triton X-100 in PBS) overnight; Alex594-conjugated anti-rabbit (Molecular Probes) (1:200) and FITC-conjugated anti-rat (Molecular Probes) (1:450) (only for ACK2 staining) in PBST for 2 h. The specimens were washed three times with PBST between steps and then mounted with slow-fade reagent (Molecular Probes). Images were captured using a Laser Scanning Confocal Image System MRC-1024 (Bio-Rad).

Transplantation

Transplantation was performed as described previously⁷. Briefly, vibrissal follicles were microdissected from ROSA 26 mice and cut into fragments that were then implanted onto the back skin of albino (OF1) neonates together with sterile black charcoal. Operated pups were observed until P5. The skin regions containing the implants, identified by the presence of charcoal, were collected and transplanted onto the back of the athymic (Swiss-nu/nu) mice to avoid rejection of the grafts. The formation of chimaeric follicles (ROSA 26–OF1) was examined by β -galactosidase staining of the skin.

Received 19 November 2001; accepted 27 February 2002.

- Schofield, R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* **4**, 7–25 (1978).
- Watt, F. M. & Hogan, B. L. Out of Eden: stem cells and their niches. *Science* **287**, 1427–1430 (2000).
- Xie, T. & Spradling, A. C. A niche maintaining germ line stem cells in the *Drosophila* ovary. *Science* **290**, 328–330 (2000).
- Slominski, A. & Paus, R. Melanogenesis is coupled to murine anagen: toward new concepts for the role of melanocytes and the regulation of melanogenesis in hair growth. *J. Invest. Dermatol.* **101**, 90s–97s (1993).
- Mackenzie, M. A., Jordan, S. A., Budd, P. S. & Jackson, I. J. Activation of the receptor tyrosine kinase Kit is required for the proliferation of melanoblasts in the mouse embryo. *Dev. Biol.* **192**, 99–107 (1997).
- Jordan, S. A. & Jackson, I. J. A late wave of melanoblast differentiation and rostrocaudal migration revealed in patch and rump-white embryos. *Mech. Dev.* **92**, 135–143 (2000).
- Kunisada, T. *et al.* Murine cutaneous mastocytosis and epidermal melanocytosis induced by keratinocyte expression of transgenic stem cell factor. *J. Exp. Med.* **187**, 1565–1573 (1998).
- Cotsarelis, G., Sun, T. T. & Lavker, R. M. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61**, 1329–1337 (1990).
- Oshima, H., Rochat, A., Kedzia, C., Kobayashi, K. & Barrandon, Y. Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell* **104**, 233–245 (2001).
- Lavker, R. M. *et al.* Hair follicle stem cells: their location, role in hair cycle, and involvement in skin tumour formation. *J. Invest. Dermatol.* **101**, 16s–26s (1993).
- Chase, H. B. Critical stages of hair development and pigmentation in the mouse. *Physiol. Zool.* **24**, 1–8 (1951).
- Tobin, D. J., Hagen, E., Botchkarev, V. A. & Paus, R. Do hair bulb melanocytes undergo apoptosis during hair follicle regression (catagen)? *J. Invest. Dermatol.* **111**, 941–947 (1998).
- Loeffler, M. & Potten, C. S. in *Stem Cells* (ed. Potten, S. C.) 5–6 (Academic, London, 1997).
- Silver, A. F., Chase, H. B. & Potten, C. S. Melanocyte precursor cells in the hair follicle germ during the dormant stage (telogen). *Experientia* **25**, 299–301 (1968).
- Cock, A. G. & Cohen, J. The melanoblast reservoir available to a feather papilla. *Embryology* **6**, 530–545 (1958).

16. Sugiyama, S. Mode of redifferentiation and melanogenesis of melanocytes in mouse hair follicles. An ultrastructural and cytochemical study. *J. Ultrastruct. Res.* **67**, 40–54 (1979).
17. Slominski, A. *et al.* Pharmacological disruption of hair follicle pigmentation by cyclophosphamide as a model for studying the melanocyte response to and recovery from cytotoxic drug damage *in situ*. *J. Invest. Dermatol.* **106**, 1203–1211 (1996).
18. Botchkareva, N. V., Khlgatian, M., Longley, B. J., Botchkarev, V. A. & Gilchrist, B. A. SCF/c-kit signalling is required for cyclic regeneration of the hair pigmentation unit. *FASEB J.* **15**, 645–658 (2001).
19. Nishikawa, S. *et al.* *In utero* manipulation of coat colour formation by a monoclonal anti-c-kit antibody: two distinct waves of c-kit-dependency during melanocyte development. *EMBO J.* **10**, 2111–2118 (1991).
20. Okura, M., Maeda, H., Nishikawa, S. & Mizoguchi, M. Effects of monoclonal anti-c-kit antibody (ACK2) on melanocytes in newborn mice. *J. Invest. Dermatol.* **105**, 322–328 (1995).
21. Ogawa, M. *et al.* Expression and function of c-Kit in fetal hemopoietic progenitor cells: transition from the early c-Kit-independent to the late c-Kit-dependent wave of hemopoiesis in the murine embryo. *Development* **117**, 1089–1098 (1993).
22. Yoshinaga, K. *et al.* Role of c-kit in mouse spermatogenesis: identification of spermatogonia as a specific site of c-kit expression and function. *Development* **113**, 689–699 (1991).
23. Mayer, T. C. The migratory pathway of neural crest cells into the skin of mouse embryos. *Dev. Biol.* **34**, 39–46 (1973).
24. Nishimura, E. K., Yoshida, H., Kunisada, T. & Nishikawa, S. I. Regulation of E- and P-cadherin expression correlated with melanocyte migration and diversification. *Dev. Biol.* **215**, 155–166 (1999).
25. Silvers, W. K. *The Coat Colours of Mice. A Model for Gene Action and Interaction* (Springer, New York, 1979).
26. Mann, S. J. Prenatal formation of hair follicle types. *Anat. Rec.* **144**, 135–142 (1962).
27. Friedlich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**, 1513–1523 (1991).
28. Yoshida, H. *et al.* Neural and skin cell-specific expression pattern conferred by steel factor regulatory sequence in transgenic mice. *Dev. Dyn.* **207**, 222–232 (1996).
29. Kunisada, T. *et al.* Transgene expression of steel factor in the basal layer of epidermis promotes survival, proliferation, differentiation and migration of melanocyte precursors. *Development* **125**, 2915–2923 (1998).
30. Fitzpatrick, T. B., Elsen, A. Z. & Wolff, K. Vitiligo. in *Dermatology in General Medicine* 3rd edn (ed. Fitzpatrick, T. B.) 810 (McGraw Hill, New York, 1987).

Acknowledgements

We thank G. Motyckova, R. Yu and A. M. Kenny for critical reading of the manuscript, and E. Nishioka for scientific support. This study was partly supported by grants from the Japanese Ministry of Education, Science and Culture. Work in the MRC Human Genetics Unit is supported by the UK Medical Research Council. E.K.N. was supported by a Research Fellowship from the Japan Society for the Promotion of Science.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.-L.N. (e-mail: snishika@virus.kyoto-u.ac.jp).

Protein kinase C δ controls self-antigen-induced B-cell tolerance

Ingrid Mecklenbräuer*, Kaoru Saijo*, Nai-Ying Zheng*, Michael Leitges† & Alexander Tarakhovsky*

* Laboratory of Lymphocyte Signaling, The Rockefeller University, New York, New York 10021, USA

† Max Planck Institute for Experimental Endocrinology, D-30625 Hannover, Germany

Interaction of a B cell expressing self-specific B-cell antigen receptor (BCR) with an auto-antigen results in either clonal deletion or functional inactivation^{1–3}. Both of these processes lead to B-cell tolerance and are essential for the prevention of auto-immune diseases. Whereas clonal deletion results in the death of developing autoreactive B cells, functional inactivation of self-reactive B lymphocytes leads to complex changes in the phenotype of peripheral B cells, described collectively as anergy^{1–3}. Here we demonstrate that deficiency in protein kinase C δ (PKC- δ) prevents B-cell tolerance, and allows maturation and terminal differentiation of self-reactive B cells in the presence of

the tolerizing antigen. The importance of PKC- δ in B-cell tolerance is further underscored by the appearance of autoreactive anti-DNA and anti-nuclear antibodies in the serum of PKC- δ -deficient mice. As deficiency of PKC- δ does not affect BCR-mediated B-cell activation *in vitro* and *in vivo*, our data suggest a selective and essential role of PKC- δ in tolerogenic, but not immunogenic, B-cell responses.

Protein kinase C δ (PKC- δ) belongs to the subfamily of PKCs that include PKC- δ , - ϵ , - η and - θ (ref. 4). In lymphocytes, PKC- δ is expressed at highest levels in developing pro- and pre-B cells, and at lowest levels in mature T cells (Supplementary Information Fig. 1a, b). Incubation of splenic B cells with anti-immunoglobulin- μ (IgM) antibody increases phosphorylation of PKC- δ at threonine 505, located within the activation loop of PKC- δ (ref. 5; see also Supplementary Information Fig. 1c). Involvement of PKC- δ in BCR signalling and the ability of PKC- δ to regulate growth, apoptosis and differentiation of cells of various types^{6–9} points towards a potentially important role of PKC- δ in antigen-dependent control of B-cell function.

The role of PKC- δ in B-cell immunity was addressed by the analysis of mice homozygous for the null mutation in the *PKC- δ* gene¹⁰. The absence of full-length or truncated PKC- δ in lymphocytes was confirmed by immunoblot analysis of cell lysates (Supplementary Information Fig. 1b). Deficiency in PKC- δ does not induce changes in the expression levels of PKC- ϵ , - η or - μ (Supplementary Information Fig. 2). The latter result excludes an indirect effect of the PKC- δ deficiency on functional properties of the

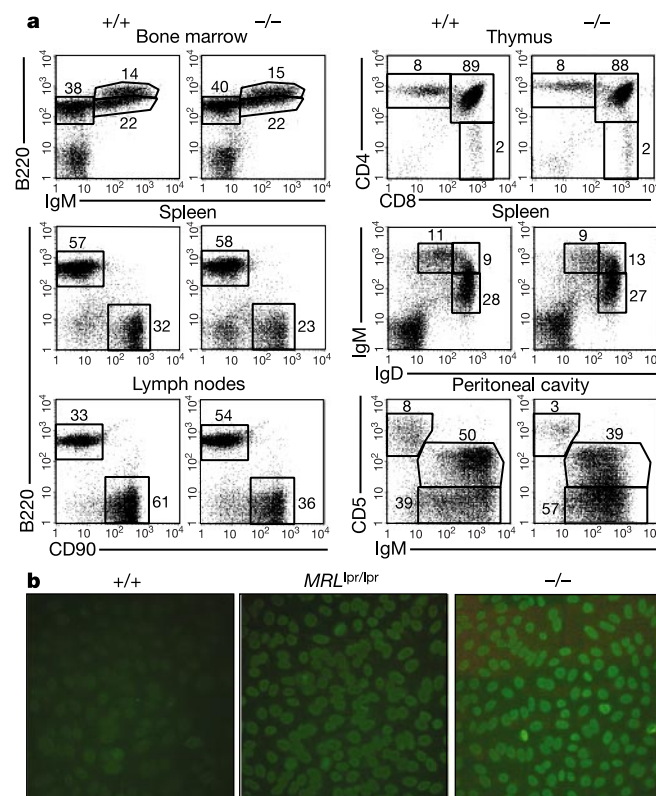


Figure 1 Lymphocyte development and auto-antibody production in the absence of PKC- δ . **a**, Results of representative FACS analysis of lymphocyte populations of eight-week-old PKC- δ ^{-/-} and PKC- δ ^{+/+} littermates are shown. Numbers indicate the percentage of cells within the gated lymphocyte subpopulation. **b**, Anti-nuclear antibodies in the serum of PKC- δ ^{-/-} mice. Sera derived from nine-month-old PKC- δ ^{+/+} (+/+), PKC- δ ^{-/-} (-/-), or four-month-old autoimmune MRL^{lpr/lpr} mice, at a dilution of 1:200, were used for the immunofluorescent analysis of anti-nuclear antibodies. Bound IgG was detected with anti-mouse IgG-FITC-conjugated antibody.

PKC- δ ^{-/-} B cells through changes in the expression of other PKCs of the same subfamily or PKC- μ .

The development of B- and T-lineage cells is not affected by the absence of PKC- δ (Fig. 1a). However, the spleens and lymph nodes of the mutant (*PKC- δ* ^{-/-}) mice are enlarged (Supplementary Information Fig. 3a). The splenomegaly and lymphadenopathy are associated with up to a twofold and sevenfold increase numbers of B cells, respectively, as well as a twofold increase in the number of erythroid (TER119⁺) cells in the spleen of mutant mice. The numbers of B1a (CD5⁺ IgM⁺), myeloid (Mac-1⁺ and Gr1⁺), natural killer (DX5⁺ CD3e⁻) and dendritic (CD11c⁺) cells in the spleens of *PKC- δ* ^{-/-} mice are unaltered compared to wild-type mice (data not shown).

An increase in the number of B cells in the lymphoid organs of mutant mice correlates with a mild titre increase of IgM, IgG1 and IgA in the serum of *PKC- δ* ^{-/-} mice, as compared with wild-type control mice (Supplementary Information Fig. 3b). The serum of *PKC- δ* ^{-/-} mice contains relatively high amounts of anti-DNA IgG1 antibodies, as well as polyreactive antibodies that recognize antigens such as nitrophenyl or chicken γ -globulin, without previous immunization with these antigens (Supplementary Information Fig. 3c). Notably, the serum of *PKC- δ* ^{-/-} mice ($n = 10$) at 7–9 months of age contains anti-nuclear antibodies (Fig. 1b). Production of autoreactive antibodies suggests abnormal selection of autoreactive B cells in the absence of PKC- δ , and may contribute towards the premature death of *PKC- δ* ^{-/-} mice observed at 9–12 months of age.

To determine whether PKC- δ is involved in negative selection of self-reactive B cells, we used immunoglobulin-transgenic mice that bear the hen egg lysozyme-specific immunoglobulin receptor (IgHEL)¹. Breeding of IgHEL mice to transgenic mice that express membrane-bound or soluble forms of HEL (mHEL or sHEL, respectively) generates double-transgenic mice in which HEL has the function of the auto-antigen^{1,3}. In IgHEL; mHEL mice, HEL-binding B cells are deleted in the bone marrow³. In contrast, the HEL-binding B cells can develop in IgHEL; sHEL mice, but become functionally incapacitated (anergized)¹.

The expression levels of HEL-binding surface immunoglobulins are similar in B cells derived from the bone marrow and spleens of *PKC- δ* ^{+/+} IgHEL and *PKC- δ* ^{-/-} IgHEL mice (Fig. 2a). The number of the IgHEL-positive B-lineage (B220⁺) cells are similar in the bone marrow of *PKC- δ* ^{+/+} IgHEL and *PKC- δ* ^{-/-} IgHEL mice (Supplementary Information Fig. 4a). An encounter of developing IgHEL-expressing B cells with mHEL in the bone marrow of *PKC- δ* ^{+/+} and *PKC- δ* ^{-/-} mice leads to the clonal deletion of newly generated B cells (data not shown). This result shows that PKC- δ is not essential for the clonal deletion induced by membrane-associated self antigens.

Interaction of HEL-binding B cells from *PKC- δ* ^{+/+} mice with sHEL in the bone marrow leads to a 7–11-fold reduction in the surface expression levels of IgMa (Fig. 2a). In contrast, the expression levels of IgMa on bone-marrow-derived B cells as well as splenic B cells in *PKC- δ* ^{-/-} IgHEL; sHEL mice are reduced by twofold compared with *PKC- δ* ^{-/-} IgHEL mice (Fig. 2a). These

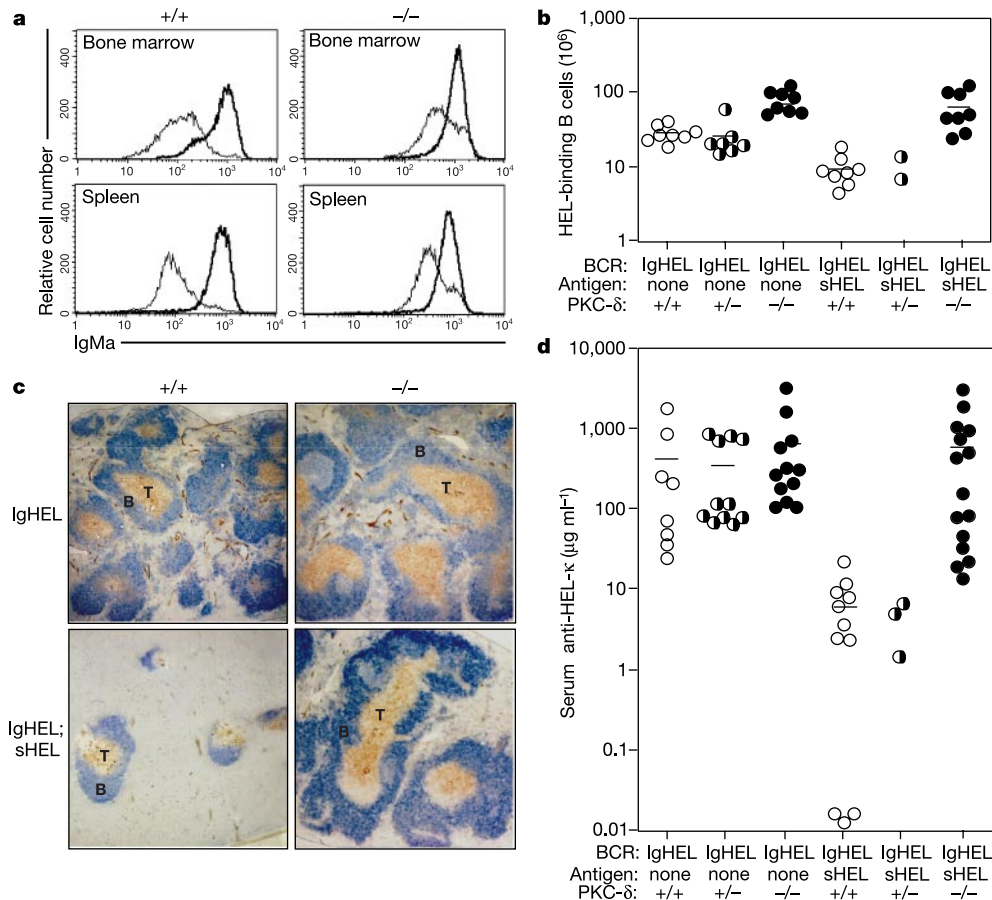


Figure 2 Defective anergy induction in the absence of PKC- δ . **a**, Histograms showing expression levels of surface IgMa on the bone marrow (upper panel) or splenic (lower panel) B cells of *PKC- δ* ^{+/+} (left column) or *PKC- δ* ^{-/-} (right column) IgHEL (thick line) or IgHEL; sHEL (thin line) mice. **b**, Number of HEL-binding B cells in the spleen. Each dot represents an individual mouse. **c**, Lymphoid follicles in the spleen. Cryosections were

stained with anti-B220 (blue, B) and anti-CD3e (red, T). Pictures were taken at $\times 5$ magnification. **d**, Anti-HEL antibody in the serum of *PKC- δ* ^{-/-} (open circles), *PKC- δ* ^{+/+} (half-filled circles) and *PKC- δ* ^{-/-} (filled circles) mice carrying IgHEL alone or in combination with sHEL. Horizontal bars in **b** and **d** represent mean values.

results show the positive role of PKC- δ in signalling leading to the downregulation of surface IgM caused by chronic exposure to sHEL. It also suggests that the PKC- δ -deficient B cells might be refractory to tolerance-inducing signals generated during B-cell exposure to the self antigen.

Chronic exposure of PKC- $\delta^{+/+}$ or PKC- $\delta^{-/-}$ IgHEL B cells to sHEL leads to a reduction in the number of peripheral B cells in IgHEL; sHEL mice compared with IgHEL mice (Fig. 2b). The absence of PKC- δ averts the negative effect of sHEL on IgHEL B-cell accumulation, as evidenced by the equal number of HEL-binding cells in the spleen and lymph nodes of PKC- $\delta^{-/-}$ IgHEL and IgHEL; sHEL mice (Fig. 2b; and data not shown). Chronic encounter of sHEL by HEL-binding cells leads to changes in the B-cell phenotype characteristic for the tolerant status of these cells^{11,12}. Tolerant B cells are characterized by low and high surface expression levels of CD23 and CD5, respectively (Supplementary Information Fig. 4b). In contrast, a significant fraction of splenic and lymph node B cells in PKC- $\delta^{-/-}$ IgHEL; sHEL mice consists of CD23⁺ CD5⁺ cells (Supplementary Information Fig. 4b). The positive role of sHEL in maturation of PKC- $\delta^{-/-}$ autoreactive B cells in IgHEL; sHEL mice was confirmed by histological analyses of cryosections of spleens derived from PKC- $\delta^{+/+}$ and PKC- $\delta^{-/-}$ IgHEL and IgHEL; sHEL mice. Whereas the spleens of PKC- $\delta^{+/+}$ IgHEL; sHEL mice have reduced numbers of B-cell follicles, the PKC- δ -deficient B cells generate large numbers of these follicles in IgHEL; sHEL mice (Fig. 2c). Defective tolerance induction in the absence of PKC- δ is supported further by the 80-fold increase in the titres of HEL-specific antibodies in the serum of PKC- $\delta^{-/-}$ IgHEL; sHEL mice compared with the serum of PKC- $\delta^{+/+}$ and PKC- $\delta^{+/+}$ IgHEL; sHEL mice (Fig. 2d). These results demonstrate that HEL-binding, self-reactive B cells deficient for PKC- δ do not become tolerant, but continue to mature, accumulate and differentiate into antibody-producing cells.

The efficiency of sHEL-induced tolerance induction is sensitive to variations in the expression level of sHEL and the occupancy of IgHEL on B cells¹³. Levels of sHEL in the serum of PKC- $\delta^{-/-}$ IgHEL or IgHEL; sHEL mice are similar to those of PKC- $\delta^{+/+}$ controls (Supplementary Information Fig. 5a). Furthermore, the amount of

endogenously produced sHEL bound to surface IgHEL on splenic B cells derived from PKC- $\delta^{-/-}$ IgHEL; sHEL mice is higher than on B cells of PKC- $\delta^{+/+}$ IgHEL; sHEL mice (Supplementary Information Fig. 5b). To measure the degree of receptor occupancy by sHEL on wild-type and mutant B cells, we compared the amount of endogenously produced sHEL bound to *ex vivo* isolated B cells, to the amount of exogenous HEL bound to IgHEL B cells upon their *in vitro* incubation with a saturating (500 ng ml⁻¹) amount of antigen. As indicated by this analysis, the amount of endogenous sHEL bound to the surface of PKC- $\delta^{-/-}$ IgHEL and PKC- $\delta^{+/+}$ IgHEL B cells is close to saturation (Supplementary Information Fig. 5b). Hence the observed break in tolerance is not due to differences in the level of self antigen or amount of IgHEL occupied by sHEL.

To exclude the possibility that the defective tolerance induction in PKC- $\delta^{-/-}$ IgHEL; sHEL mice may reflect a general signalling defect in PKC- $\delta^{-/-}$ mice, we analysed the responses of PKC- $\delta^{-/-}$ B cells to antigenic stimulation both *in vitro* and *in vivo*. Splenic B cells isolated from non-immunoglobulin-transgenic mice and IgHEL PKC- $\delta^{-/-}$ mice are characterized by wild-type levels of *in vitro* responses to antigenic stimulation (Fig. 3). Engagement of the BCR *in vitro* upregulates cell-surface-expressed CD86 (B7-2), and induces proliferation of PKC- $\delta^{-/-}$ B cells (Fig. 3, upper and lower panels). Furthermore, the amplitudes and kinetics of BCR-mediated calcium mobilization are similar in B cells of PKC- $\delta^{-/-}$ and PKC- $\delta^{+/+}$ mice (Fig. 3, middle panel). B cells of both PKC- $\delta^{+/+}$ and PKC- $\delta^{-/-}$ mice are prone to spontaneous death *in vitro*, and are equally sensitive to apoptosis induced by anti-Fas or anti-IgM *in vitro* (Supplementary Information Fig. 6).

Similar to B cells derived from non-immunoglobulin-transgenic PKC- $\delta^{-/-}$ mice, the B cells collected from spleens of PKC- $\delta^{-/-}$ IgHEL mice respond as vigorously to antigenic stimulation as B cells of PKC- $\delta^{+/+}$ IgHEL mice (Fig. 3). Deficiency in PKC- δ does not alter the threshold for HEL-induced B-cell activation, as the B cells of PKC- $\delta^{+/+}$ and PKC- $\delta^{-/-}$ mice show equally strong responses to HEL at different concentrations (Supplementary Information Fig. 7a). Furthermore, B cells of PKC- $\delta^{-/-}$ mice were able to mount wild-type-like humoral immune responses to T-cell-dependent or -independent antigens nitrophenyl-chicken γ -globulin and

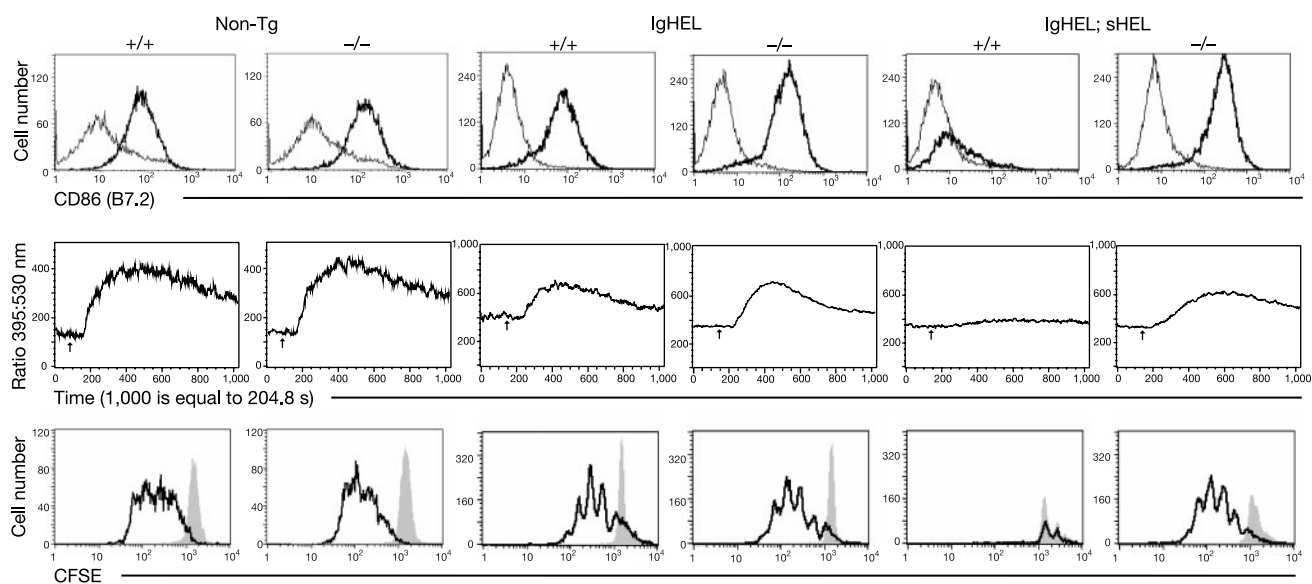


Figure 3 Activation of B cells *in vitro*. The top panel shows CD86 upregulation in splenic B cells incubated in the presence (thick line) or absence (thin line) of 2.5 µg ml⁻¹ anti-IgM (non-Tg) or with 10 ng ml⁻¹ HEL (IgHEL and IgHEL; sHEL) for 24 h. The middle panel shows BCR-mediated calcium mobilization. Changes in the fluorescence of Indo-1-loaded B cells was monitored before and after stimulation (indicated by an arrow) and are shown

as lines representing the mean value of the response. The bottom panel shows BCR-mediated B-cell proliferation. CFSE fluorescence of unstimulated B cells (grey area) or B cells stimulated with anti-IgM in combination with IL-4 (non-Tg), or 1,000 ng ml⁻¹ HEL (IgHEL and IgHEL; sHEL) for 3 days is shown.

nitrophenyl-Ficoll, respectively (data not shown). These data show that PKC- δ deficiency does not interfere with immunogenic signals leading to B-cell activation *in vitro* and *in vivo*.

Despite unaltered antigen-induced activation of B cells from PKC- $\delta^{-/-}$ mice, the analysis of *in vitro* responses of B cells derived from PKC- $\delta^{-/-}$ IgHEL; sHEL mice revealed that chronic exposure to sHEL fails to induce tolerance in B cells (Fig. 3). Whereas incubation of PKC- $\delta^{+/+}$ tolerant B cells with HEL does not upregulate CD86 (Fig. 3, upper panel), similar treatment of B cells derived from PKC- $\delta^{-/-}$ IgHEL; sHEL mice leads to the upregulation of CD86 (Fig. 3, upper panel; see also Supplementary Information Fig. 7a). Incubation of tolerant B cells with HEL does not induce their proliferation *in vitro* (Fig. 3, lower panel). By contrast, B cells derived from PKC- $\delta^{-/-}$ IgHEL; sHEL mice proliferate vigorously in response to HEL (Fig. 3, lower panel). Finally, incubation with HEL fails to induce calcium mobilization in tolerant B cells of PKC- $\delta^{+/+}$ mice (Fig. 3, middle panel). In contrast to tolerant B cells, the magnitude of calcium mobilization in HEL-

triggered B cells derived from PKC- $\delta^{-/-}$ IgHEL; sHEL mice is close to wild-type levels (Fig. 3, middle panel). Collectively, these data show that the B cells of IgHEL PKC- $\delta^{-/-}$ mice that develop in the presence of the tolerogen (sHEL) do not manifest any of the principal signalling/functional alterations associated with B-cell tolerance within the experimental system.

Tolerant B cells are defective in their responses not only to antigen, but also to antibody mediated cross-linking of B-cell-expressed receptor proteins RP105 and CD38 (Supplementary Information Fig. 7b). Defective signalling through BCR, RP105 and CD38 in tolerant B cells is rather specific, as *in vitro* responses of these cells to the unmethylated oligonucleotide (CpG) and the anti-complement receptor (anti-CD21/CD35) antibody are unaltered (Supplementary Information Fig. 7b). Deficiency in PKC- δ restores completely the CD38- and RP105-mediated activation of B cells derived from IgHEL; sHEL mice (Supplementary Information Fig. 7b). This result shows that deficiency in PKC- δ prevents the development of tolerance-specific signalling defects not only down-

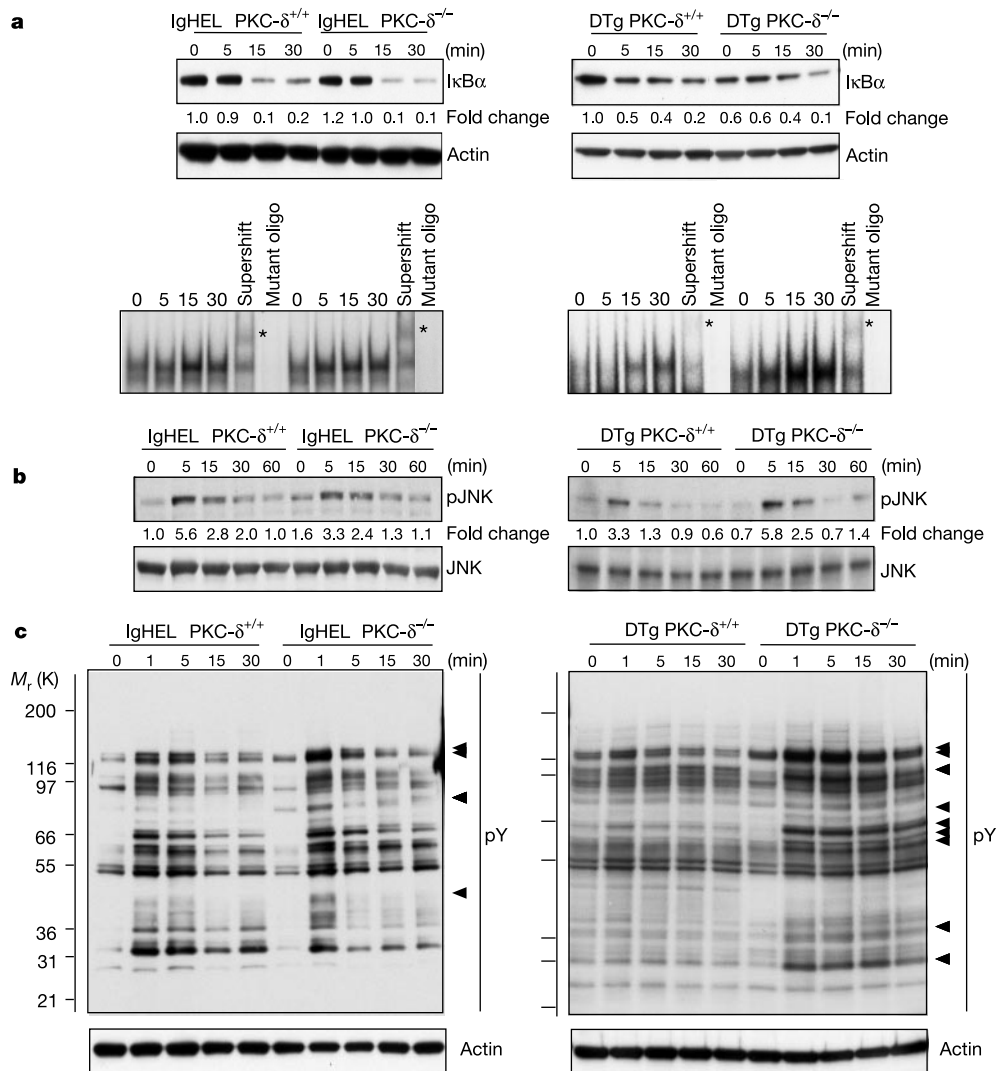


Figure 4 BCR-mediated signalling in B cells of PKC- $\delta^{-/-}$ mice. Purified B cells from IgHEL PKC- $\delta^{+/+}$ and IgHEL PKC- $\delta^{-/-}$ mice (left panels) or IgHEL; sHEL mice (double transgenic (DTg) PKC- $\delta^{+/+}$ and PKC- $\delta^{-/-}$; right panels) were stimulated with 500 ng ml $^{-1}$ HEL. **a**, The expression of IkB α (top panel) was analysed by immunoblotting with anti-IkB α antibody. NF- κ B DNA-binding activity was examined by electrophoresis mobility shift assay (lower panel). Oligo, oligonucleotide. **b**, c, Phosphorylation of JNK (**b**) and total protein tyrosine phosphorylation (**c**) was analysed by immunoblotting with an

anti-phospho-JNK-specific or anti-phosphotyrosine antibody (pY), respectively. Arrowheads in **c** indicate the protein bands of altered signal intensity. Numbers show the fold change of the intensity of the protein band signal normalized for the intensity of the α -actin signal, and relative to unstimulated PKC- $\delta^{+/+}$ IgHEL or IgHEL; sHEL B cells, which was set to 1.0. Equal protein loading was confirmed by re-probing immunoblots with anti- α -actin (**a**, **c**) or anti-JNK (**b**) antibody. Asterisks indicate the supershifted c-Rel and p50-containing complex.

stream of the BCR, but also of other receptor proteins. A pronounced break in B-cell tolerance was also found in mice deficient for complement receptor¹². However, unaltered responses of *PKC-δ*^{-/-} B cells to antibody mediated cross-linking of the complement receptor (CD21/35) *in vitro* (Supplementary Information Fig. 7b) do not support an important role for *PKC-δ* in complement receptor signalling.

Induction of tolerance in IgHEL B cells is associated with inefficient degradation of cytoplasmic IκB and reduced translocation of nuclear factor (NF)-κB to the nucleus of tolerant B cells¹⁴. Engagement of BCR on B cells of *PKC-δ*^{+/+} and *PKC-δ*^{-/-} non-transgenic and IgHEL mice results in an equally strong reduction of IκBα in the cytosol, followed by an increase of NF-κB in the nuclei (Fig. 4a, left panel; see also Supplementary Information Fig. 8a). However, the pattern of the NF-κB induction is different between tolerant and non-tolerant B cells isolated from *PKC-δ*^{+/+} and *PKC-δ*^{-/-} IgHEL; sHEL mice, respectively (Fig. 4a, top right panel). Both the basal level and HEL-mediated induction of NF-κB are significantly stronger in the nuclei of B cells isolated from *PKC-δ*^{-/-} IgHEL; sHEL mice than of B cells derived from *PKC-δ*^{+/+} IgHEL; sHEL mice (Fig. 4a, bottom right panel). Comparison of B cells from *PKC-δ*^{-/-} IgHEL and IgHEL; sHEL mice suggest differences in the mechanism of NF-κB induction between *PKC-δ*^{-/-} B cells, which develop in the absence or presence of the tolerizing antigen (Fig. 4a). The slower and less efficient IκBα degradation in B cells derived from *PKC-δ*^{-/-} IgHEL; sHEL mice suggests that NF-κB induction in these cells is dependent on the reduction of IκBα to a lesser extent than in B cells of *PKC-δ*^{-/-} IgHEL mice. The *PKC-δ*^{-/-} IgHEL B cells are characterized by modest but constantly observed higher basal levels of Jun N-terminal kinase (JNK) phosphorylation, as compared with B cells from *PKC-δ*^{+/+} IgHEL mice. In both cases, engagement of the BCR results in an increase in JNK phosphorylation (Fig. 4b, upper left). Incubation of tolerant B cells with HEL induces phosphorylation of JNK to a lesser degree as compared with the stimulated non-tolerant B cells¹⁴ (Fig. 4b, upper right panel). The induction of JNK phosphorylation in B cells isolated from *PKC-δ*^{-/-} IgHEL; sHEL mice has a pattern similar to one observed in tolerant B cells isolated from the *PKC-δ*^{+/+} IgHEL; sHEL mice. Although the primary cause for efficient NF-κB activation in the absence of *PKC-δ* remains unknown, it might be related to the altered pattern of BCR-induced tyrosine phosphorylation of several proteins (Fig. 4c; see also Supplementary Information Fig. 8b) in B cells isolated from *PKC-δ*^{-/-} IgHEL and IgHEL; sHEL mice, as compared with control B cells from *PKC-δ*^{+/+} mice (Fig. 4c). It is possible that similar to *PKC-β1/II*, which regulates negatively the activity of the protein tyrosine kinase *Btk*^{15,16}, *PKC-δ* might control the activity of an as yet unidentified tyrosine kinase/phosphatase.

Overall, the observed signalling changes yield a new signalling pattern characteristic for *PKC-δ*^{-/-} IgHEL B cells that develop in the presence of sHEL. This is particularly true for the NF-κB induction that occurs in the absence of significant IκBα degradation and is, therefore, conflicting with a canonical IκB-dependent mechanism of NF-κB activation. Nevertheless, the IκB degradation independent of NF-κB induction in IgG⁺ B cells¹⁷ supports the possibility of such a mechanism. Alteration of nuclear signalling in tolerant B cells is likely to be one of the chief causes of a global change in the pattern of gene expression in tolerant B cells¹⁸. Conversely, the restoration of nuclear signalling in the absence of *PKC-δ* should re-activate genes required for B-cell activation and differentiation.

Our findings present *PKC-δ* as an essential component of a signalling pathway specific for the induction of tolerance in B cells. It remains to be tested whether aberrations in *PKC-δ* expression and/or activity in B cells are associated with the development of auto-immune diseases in humans. □

Methods

Mice

We used *PKC-δ*^{-/-} mice on a C57Bl/6 background for the analysis¹⁰. The *PKC-δ*^{-/-} IgHEL and IgHEL; sHEL or IgHEL; mHEL transgenic mice were generated by breeding to MD4 IgHEL, KLK3 mHEL and ML5 sHEL transgenic mice (gift of C. Goodnow), and were genotyped by polymerase chain reaction of tail DNA, as described^{1,3}.

Cell staining and flow cytometry

Analysis of surface marker expression on lymphocytes has been done using monoclonal antibodies against CD4 (RM4-5), B220 (RA3-6B2), IgD (1.3-5), IgM (RS3.1), IgM (R33-24.12), Thy1.2 (Cfo-1), CD5 (53-7.3), CD8a (Ly-2), CD19 (1D3), CD23 (B3B4), CD43 (S7) and CD86 (GL1), as described previously¹⁹. Expression of β-galactosidase was examined by using fluorescein di-β-D-galactopyranoside (FDG; Molecular Probes) as described²⁰. Lymphocyte populations were analysed by fluorescence-activated cell sorting (FACS) on a FACScan (Becton Dickinson). HEL-transgenic B cells were detected by staining with HEL conjugated to phycoerythrin (PE). Dead cells were revealed by staining with TO-PRO-3 (Molecular Probes), and excluded from the analysis.

ELISA and ANA

We performed enzyme-linked immunosorbent assay (ELISA) as described¹⁵. The serum titres of lysozyme-specific κ-antibodies were measured by ELISA on microtitre plates coated with 10 μg ml⁻¹ lysozyme (Sigma). We used biotinylated goat anti-mouse κ-antibody for the detection of lysozyme-bound antibodies. The concentration of serum antibodies was determined by using anti-HEL antibody 2D1 (ref. 21) as a standard. Serum HEL titres were assayed by ELISA on microtitre plates coated with 10 μg ml⁻¹ anti-HEL antibody HyHEL10 (ref. 22). We used biotinylated anti-HEL antibody 2D1 (ref. 21) for the detection of serum HEL. The serum titres of DNA-, nitrophenyl- and chicken γ-globulin-specific IgG1 antibodies were measured by ELISA on microtitre plates coated with 10 μg ml⁻¹ single-stranded salmon sperm DNA, nitrophenyl₄-BSA, or chicken γ-globulin, and biotinylated goat anti-mouse IgG1 was used for detection. Anti-nuclear antibodies (ANA) were detected by staining HEP-2 slides (Bion Enterprises) with mouse serum at 1:200 for 20 min, followed by anti-mouse IgG-FITC conjugated antibody (Jackson ImmunoResearch) for 15 min.

B-cell proliferation and upregulation of activation markers

Splenic B cells were purified as described previously¹⁵, and stimulated *in vitro* with 2.5 μg ml⁻¹ F(ab')₂ fragment goat anti-mouse IgM (Jackson ImmunoResearch) in combination with 25 U ml⁻¹ recombinant mouse interleukin (IL)-4 (Genzyme), or 5.25 μg ml⁻¹ anti-RP105 (ref. 23), 7.5 μg ml⁻¹ anti-CD38 (ref. 24), 10 nM CpG ODN 1638 (ref. 25) (MWG Biotech), or lysozyme from chicken egg white (Sigma) at various concentrations (1–1,000 ng ml⁻¹). We carried out cross-linking of CD21/CD35 as described²⁶. The analysis of the upregulation of CD86 was performed as described previously¹⁹. Labelling of cells with 5- and 6-carboxyfluorescein diacetate, succinimidyl ester (CFDA-SE; Molecular Probes) for analysis of proliferation was performed as described²⁷. The decline in CFSE fluorescence as a measure of B-cell proliferation was determined by FACS analysis.

Calcium mobilization

Splenocytes were loaded with Indo-1 (Molecular Probes) and stained with anti-CD19-PE-conjugated antibody. Calcium mobilization was initiated by adding 20 μg ml⁻¹ anti-IgM or 200 ng ml⁻¹ HEL to the sample 20 s after the start of acquisition. Indo-1 fluorescence was recorded at 395 nm and 530 nm on a FACSort equipped with an ultraviolet laser. The 395- versus 530-nm ratio was used as a readout for changes in intracellular free calcium concentrations, and analysed using FlowJo (Tristar) software.

Immunohistochemical analysis

We froze spleens in tissue-freezing medium (Jung). Cryosections were incubated with biotinylated antibodies against B220 (RA3-6B2) and CD3ε (145-2C11) (BD Pharmingen), followed by substrates for ABC-P (Vector NovaRED) and ABC-AP (Vector Blue) from Vector Laboratories.

Immunoblotting analysis and EMSA

For the analysis of *PKC* isoform expression, the protein extracts of purified splenic B cells from *PKC-δ*^{-/-} and control mice were analysed by immunoblotting with the antibodies against *PKC-δ*, *PKC-η* (BD Transduction Laboratory), *PKC-θ* (Santa Cruz), *PKC-ε* (UBI) and *PKC-μ* (New England Biolabs). Protein loading was controlled by immunoblotting with anti-α-actin antibody (Oncogene). The expression of IκBα and NF-κB activation was studied using cytoplasmic and nuclear protein extracts prepared as described¹⁷. We measured expression of IκBα by immunoblotting with anti-IκBα antibody (Santa Cruz). We analysed DNA-binding activity of NF-κB by electrophoretic mobility shift assay (EMSA). Nuclear extracts (4 μg) were incubated for 30 min at room temperature with 2 μg poly(dI-dC) (Pharmacia) and 0.5 ng ³²P-labelled NFκB consensus or mutant oligonucleotide (Santa Cruz) in binding buffer (10 mM TrisCl, pH 7.5, 50 mM NaCl, 0.5 mM dithiothreitol, 0.1 mM EDTA, 0.1 mg ml⁻¹ BSA and 4% glycerol). The supershift assay was performed by pre-incubation of nuclear extracts with anti-c-Rel and anti-p50 antibodies (Santa Cruz), before addition of the labelled probe. Samples were fractionated on a 5% native polyacrylamide gel electrophoresis gel and visualized by autoradiography. The phosphorylation of total cellular proteins and JNK was analysed by immunoblotting

with anti-phospho-tyrosine (PY99) and anti-phospho-JNK (Thr183/Tyr185), respectively. Anti-JNK antibodies were purchased from Cell Signaling Technology. Data were quantified by National Institutes of Health Image software.

Received 27 December 2001; accepted 11 February 2002.

- Goodnow, C. C. *et al.* Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. *Nature* **334**, 676–682 (1988).
- Nemazee, D. A. & Burki, K. Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes. *Nature* **337**, 562–566 (1989).
- Hartley, S. B. *et al.* Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens. *Nature* **353**, 765–769 (1991).
- Nishizuka, Y. Intracellular signaling by hydrolysis of phospholipids and activation of protein kinase C. *Science* **258**, 607–614 (1992).
- Le Good, J. A. *et al.* Protein kinase C isotypes controlled by phosphoinositide 3-kinase through the protein kinase PDK1. *Science* **281**, 2042–2045 (1998).
- Watanabe, T. *et al.* Cell division arrest induced by phorbol ester in CHO cells overexpressing protein kinase C- δ subspecies. *Proc. Natl Acad. Sci. USA* **89**, 10159–10163 (1992).
- Mischak, H. *et al.* Overexpression of protein kinase C- δ and - ϵ in NIH 3T3 cells induces opposite effects on growth, morphology, anchorage dependence, and tumorigenicity. *J. Biol. Chem.* **268**, 6090–6096 (1993).
- Emoto, Y. *et al.* Proteolytic activation of protein kinase C delta by an ICE-like protease in apoptotic cells. *EMBO J.* **14**, 6148–6156 (1995).
- Ghayur, T. *et al.* Proteolytic activation of protein kinase C delta by an ICE/CED 3-like protease induces characteristics of apoptosis. *J. Exp. Med.* **184**, 2399–2404 (1996).
- Leites, M. *et al.* Exacerbated vein graft arteriosclerosis in protein kinase C δ -null mice. *J. Clin. Invest.* **108**, 1505–1512 (2001).
- Hippen, K. L., Tze, L. E. & Behrens, T. W. CD5 maintains tolerance in anergic B cells. *J. Exp. Med.* **191**, 883–890 (2000).
- Prodeus, A. P. *et al.* A critical role for complement in maintenance of self-tolerance. *Immunity* **9**, 721–731 (1998).
- Fulcher, D. A. *et al.* The fate of self-reactive B cells depends primarily on the degree of antigen receptor engagement and availability of T cell help. *J. Exp. Med.* **183**, 2313–2328 (1996).
- Healy, J. I. *et al.* Different nuclear signals are activated by the B cell receptor during positive versus negative signaling. *Immunity* **6**, 419–428 (1997).
- Leites, M. *et al.* Immunodeficiency in protein kinase C β -deficient mice. *Science* **273**, 788–791 (1996).
- Kang, S. W. *et al.* PKC β modulates antigen receptor signaling via regulation of Btk membrane localization. *EMBO J.* **20**, 5692–5702 (2001).
- Doerre, S. & Corley, R. B. Constitutive nuclear translocation of NF- κ B in B cells in the absence of I κ B degradation. *J. Immunol.* **163**, 269–277 (1999).
- Glynne, R. *et al.* How self-tolerance and the immunosuppressive drug FK506 prevent B-cell mitogenesis. *Nature* **403**, 672–676 (2000).
- Chan, V. W. *et al.* The molecular mechanism of B cell activation by toll-like receptor protein RP-105. *J. Exp. Med.* **188**, 93–101 (1998).
- Nolan, G. P., Fiering, S., Nicolas, J. F. & Herzenberg, L. A. Fluorescence-activated cell analysis and sorting of viable mammalian cells based on β -D-galactosidase activity after transduction of *Escherichia coli* lacZ. *Proc. Natl Acad. Sci. USA* **85**, 2603–2607 (1988).
- Metzger, D. W., Ch'ng, L. K., Miller, A. & Sercarz, E. E. The expressed lysozyme-specific B cell repertoire. I. Heterogeneity in the monoclonal anti-hen egg white lysozyme specificity repertoire, and its difference from the *in situ* repertoire. *Eur. J. Immunol.* **14**, 87–93 (1984).
- Smith-Gill, S. J. *et al.* Mapping the antigenic epitope for a monoclonal antibody against lysozyme. *J. Immunol.* **128**, 314–322 (1982).
- Miyake, K., Yamashita, Y., Hitoshi, Y., Takatsu, K. & Kimoto, M. Murine B cell proliferation and protection from apoptosis with an antibody against a 105-kD molecule: unresponsiveness of X-linked immunodeficient B cells. *J. Exp. Med.* **180**, 1217–1224 (1994).
- Yamashita, Y. *et al.* A monoclonal antibody against a murine CD38 homologue delivers a signal to B cells for prolongation of survival and protection against apoptosis *in vitro*: unresponsiveness of X-linked immunodeficient B cells. *Immunology* **85**, 248–255 (1995).
- Lipford, G. B. *et al.* CpG-containing synthetic oligonucleotides promote B and cytotoxic T cell responses to protein antigen: a new class of vaccine adjuvants. *Eur. J. Immunol.* **27**, 2340–2344 (1997).
- Kozono, Y. *et al.* Cross-linking CD21/CD35 or CD19 increases both B7-1 and B7-2 expression on murine splenic B cells. *J. Immunol.* **160**, 1565–1572 (1998).
- Lyons, A. B. & Parish, C. R. Determination of lymphocyte division by flow cytometry. *J. Immunol. Methods* **171**, 131–137 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank C. Goodnow, J. Cyster and F. Finkelman for providing reagents; members of the Laboratory for Lymphocyte Signaling in Cologne for technical assistance; and G. Hannon for the preparation of the manuscript. We thank M. Nussenzweig, D. O'Carroll, K. Rajewsky and C. Schmedt for discussion. This work was supported by The Irene Diamond Fund (A.T.) and National Institutes of Health grant (N.Y.Z. and A.T.), S. L. E. Foundation (K.S.), Graduiertenkolleg from the Deutsche Forschungsgemeinschaft, and The Rockefeller University's Women & Science Fellowship Program (I.M.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.T. (e-mail: tarakho@mail.rockefeller.edu).

Increased proliferation of B cells and auto-immunity in mice lacking protein kinase C δ

Akitomo Miyamoto^{*†}, Keiko Nakayama^{†‡}, Hiroyuki Imaki^{*†}, Sachiko Hirose^{†§}, Yi Jiang[§], Masaaki Abe[§], Tadasuke Tsukiyama^{*†}, Hiroyasu Nagahama^{*†}, Shigeo Ohno^{||}, Shigetsugu Hatakeyama^{*†} & Keiichi I. Nakayama^{*†‡}

^{*} Department of Molecular and Cellular Biology, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan

[†] CREST, Japan Science and Technology Corporation, Kawaguchi, Saitama 332-0012, Japan

[‡] Department of Molecular Genetics, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan

[§] Second Department of Pathology, Juntendo University School of Medicine, Bunkyo-ku, Tokyo 113-8421, Japan

^{||} Department of Molecular Biology, Yokohama City University School of Medicine, Yokohama, Kanagawa 236-0004, Japan

Protein kinase C (PKC), which comprises 11 closely related isoforms, has been implicated in a wide variety of cellular processes, such as growth, differentiation, secretion, apoptosis and tumour development^{1–4}. Among the PKC isotypes, PKC- δ is unique in that its overexpression results in inhibition of cell growth^{5–11}. Here we show that mice that lack PKC- δ exhibit expansion of the B-lymphocyte population with the formation of numerous germinal centres in the absence of stimulation. The rate of proliferation in response to stimulation was greater for B cells from PKC- δ -deficient mice than for those from wild-type mice. Adoptive transfer experiments suggested that the hyperproliferation phenotype is B-cell autonomous. Production of interleukin-6 was markedly increased in B cells of PKC- δ -null mice as a result of an increase in the DNA-binding activity of NF-IL6. Furthermore, the PKC- δ -deficient mice contain circulating autoreactive antibodies and display immune-complex-type glomerulonephritis, as well as lymphocyte infiltration in many organs. These results suggest that PKC- δ has an indispensable function in negative regulation of B-cell proliferation, and is particularly important for the establishment of B-cell tolerance.

PKC- δ is the most intensively studied member of the novel PKC subfamily⁶. Although PKC- δ is implicated in the inhibition of cell growth^{7–11}, as well as cell differentiation^{12–14}, apoptosis^{15–17} and tumour suppression^{18–22}, the role of PKC- δ in these specific cellular actions remains unclear. Here we have generated mice deficient in this PKC isozyme. The PKC- δ gene was disrupted in mouse embryonic stem cells by replacement of the first and second exons with a neomycin-resistance cassette (Fig. 1a). Neither full-length nor truncated PKC- δ protein was detected in PKC- δ ^{−/−} mouse cells by immunoblot analysis (Fig. 1b). The abundance of other PKC isoforms was similar in B lymphocytes and in the brain (see Fig. 1 of Supplementary Information) of PKC- δ ^{+/+} and PKC- δ ^{−/−} mice. PKC- δ protein was most abundant in lymphoid organs (thymus, spleen and lymph nodes) in addition to the cerebrum and intestine of wild-type mice (Fig. 1c). Both B and T lymphocytes expressed PKC- δ protein (data not shown). The PKC- δ ^{−/−} mice were viable up to 12 months of age, despite detection of auto-immune disease in these animals (see below). We have not observed cancer-induced death in either PKC- δ ^{−/−} mice ($n = 10$) or in control animals ($n = 12$) up to 12 months of age, suggesting that the lack of functional PKC- δ does not result in a predisposition towards cancer.

Necropsy revealed that the spleen and lymph nodes of PKC- δ ^{−/−}

mice were enlarged (data not shown). Flow cytometric analysis showed that the percentage of B220⁺ immunoglobulin- μ (IgM)⁺ T-cell antigen receptor (TCR) β ⁺ B lymphocytes (B-2 cells) was significantly increased in the spleen and lymph nodes of *PKC- δ ^{-/-}* mice compared with those in wild-type mice (Fig. 2a, b). Furthermore, the absolute number of B-2 cells was also increased in spleen ($49.9 \pm 16.3 \times 10^6$ in *PKC- δ ^{+/+}* mice compared with $71.5 \pm 25.2 \times 10^6$ in *PKC- δ ^{-/-}* mice, $P < 0.01$) and lymph nodes ($8.2 \pm 4.5 \times 10^6$ in *PKC- δ ^{+/+}* mice compared with $21.3 \pm 11.6 \times 10^6$ in *PKC- δ ^{-/-}* mice, $P < 0.001$). The numbers of splenic and peritoneal CD5⁺ B lymphocytes (B-1a cells) of the mutant mice were unchanged (data not shown). Both the splenomegaly and B-cell expansion were apparent in almost all *PKC- δ ^{-/-}* mice as early as 8 weeks after birth. Similar analysis of bone marrow cells did not reveal any significant differences between *PKC- δ ^{+/+}* and *PKC- δ ^{-/-}* mice (data not shown), suggesting that the expansion of the B-cell population occurs in the peripheral lymphoid tissues rather than during development. The absolute numbers of T lymphocytes in the spleen, lymph nodes and thymus of *PKC- δ ^{-/-}* mice were also essentially identical to those in wild-type mice. The CD4/CD8 and TCR α - and β -chain (TCR $\alpha\beta$) profiles of thymocytes from the *PKC- δ ^{-/-}* mice were within the normal range (data not shown). Histological analysis of the spleen of *PKC- δ ^{+/+}* mice maintained in a specific pathogen-free animal facility revealed no or a few poorly formed germinal centres in B-cell follicles (Fig. 2c). In contrast, the spleens of *PKC- δ ^{-/-}* littermates manifested numerous germinal centres, which were stained by peanut agglutinin (PNA) (Fig. 2d). Similar changes were apparent in the B-cell zones of lymph nodes (data not shown) and Peyer's patches (Fig. 2e, f).

To examine whether the B-cell phenotype of *PKC- δ ^{-/-}* mice is B-cell autonomous, we performed adoptive transfer experiments. B cells were prepared from the spleens of *PKC- δ ^{+/+}* and *PKC- δ ^{-/-}* mice, and then transferred to *Rag1^{-/-}* mice, which are virtually devoid of mature B cells. Eighteen days after cell transfer, both the percentage and the absolute number of splenic B cells were greater in mice that received *PKC- δ ^{-/-}* B cells than in those that received *PKC- δ ^{+/+}* B cells (Fig. 2g, h). Immunohistochemical analysis also revealed that germinal centre formation was markedly greater in the

spleen of mice that received *PKC- δ ^{-/-}* B cells compared with those that received *PKC- δ ^{+/+}* B cells (data not shown). These data suggested that both the B-cell expansion and the enhanced formation of germinal centres apparent in *PKC- δ ^{-/-}* mice are B-cell autonomous.

The increase in the number of re-circulating B cells in *PKC- δ ^{-/-}* mice was suggestive of an abnormality in the proliferation of these cells. We therefore measured the proliferative potential of B cells derived from the spleens of wild-type and *PKC- δ ^{-/-}* mice. Stimu-

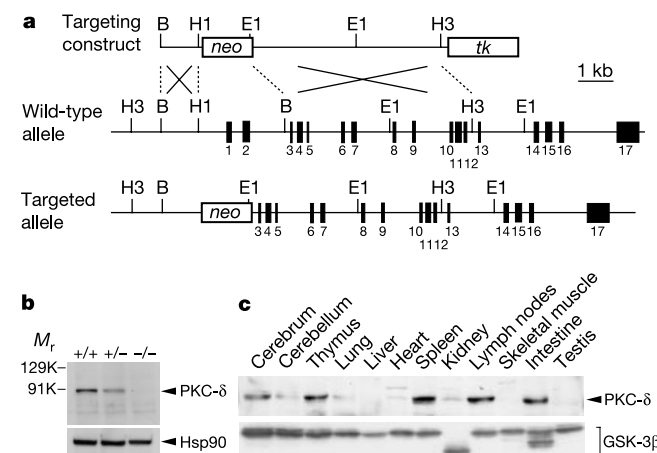


Figure 1 Targeted disruption of the mouse *PKC- δ* gene. **a**, The structures of the targeting vector, the *PKC- δ* gene and the predicted mutant locus. The exons are depicted by filled boxes. Restriction sites: B, *Bam*HI; E1, *Eco*RI; H1, *Hpa*I; H3, *Hind*III. *neo*, neomycin-resistance cassette; *tk*, thymidine kinase gene. **b**, Immunoblot analysis of *PKC- δ* and of the control protein Hsp90 in brain derived from *PKC- δ ^{+/+}*, *PKC- δ ^{+/-}* and *PKC- δ ^{-/-}* mice. **c**, Distribution of *PKC- δ* in organs of wild-type mice. Homogenates of the indicated organs were subjected to immunoblot analysis with antibodies to *PKC- δ* or to the control protein GSK-3 β .

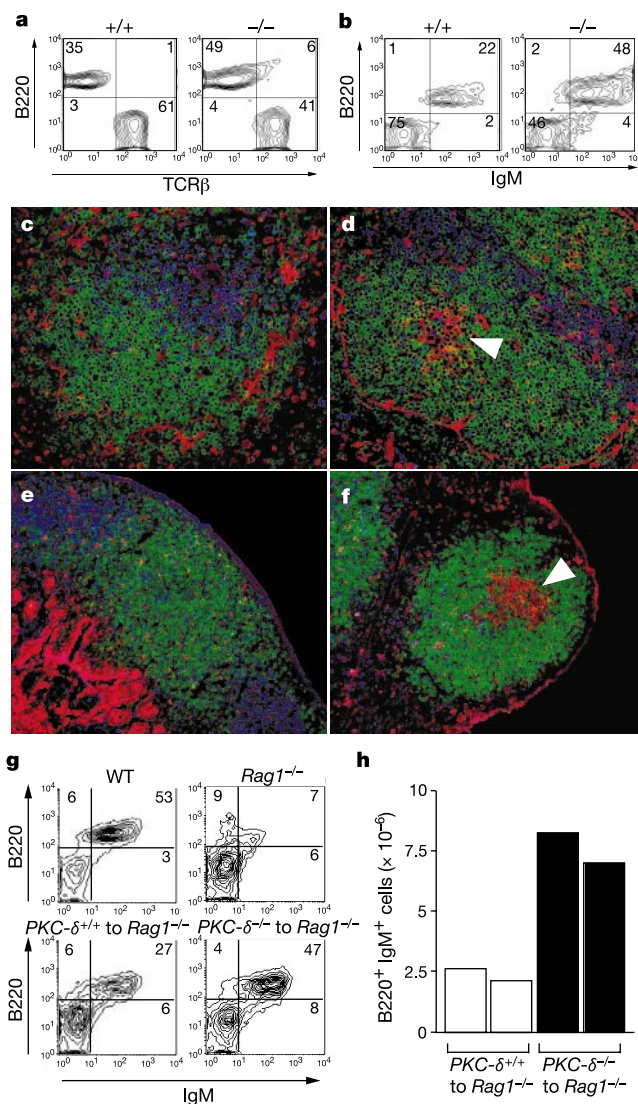


Figure 2 Expansion of the B-lymphocyte population in *PKC- δ ^{-/-}* mice. **a, b**, Flow cytometric analysis of lymphocytes derived from the lymph nodes of *PKC- δ ^{+/+}* and *PKC- δ ^{-/-}* mice, with antibodies to B220 and to TCR β (**a**) or B220 and to IgM (**b**). Percentages are shown in the corners of each graph. **c-f**, Immunofluorescence staining of sections from the spleen (**c, d**) and Peyer's patches (**e, f**) of wild-type (**c, e**) and *PKC- δ ^{-/-}* (**d, f**) mice, with antibodies to B220 (green), CD4 and CD8 (blue), and with PNA (red). Arrowheads indicate germinal centres. Original magnification: $\times 200$ (**c, d**), $\times 100$ (**e, f**). **g, h**, Analysis of *Rag1^{-/-}* mice 18 days after the adoptive transfer of B cells from *PKC- δ ^{+/+}* (*PKC- δ ^{+/+}* to *Rag1^{-/-}*) or *PKC- δ ^{-/-}* (*PKC- δ ^{-/-}* to *Rag1^{-/-}*) mice. Flow cytometric analysis of splenocytes derived from the treated animals was performed with antibodies to B220 and to IgM (**g**). Splenocytes prepared from *PKC- δ ^{+/+}* (WT) and untreated *Rag1^{-/-}* mice were also analysed. Absolute numbers of B cells in the spleen of *Rag1^{-/-}* mice that had been injected with B cells from *PKC- δ ^{+/+}* (open columns) or *PKC- δ ^{-/-}* (filled columns) mice are shown (**h**).

lation of cultures with either the combination of F(ab')₂ specific for IgM and antibodies to CD40 or with lipopolysaccharide (LPS) revealed that the proliferative response of PKC- δ -deficient B cells was markedly greater than that of B cells from wild-type mice, as evidenced by both the XTT colorimetric assay (Fig. 3a) and determination of [³H]thymidine incorporation (Fig. 3b). With such stimuli, the proportion of cells in the S phase was markedly greater for PKC- δ ^{-/-} B cells than for the wild-type cells (see Fig. 2 of Supplementary Information). Neither the extent of spontaneous

death in culture, death during mitogenic activation, and that of apoptosis induced by antibodies to Fas by etoposide or by dexamethasone differed substantially between B cells of PKC- δ ^{+/-} and PKC- δ ^{-/-} mice (see Fig. 3 of Supplementary Information). The growth properties of T cells from PKC- δ ^{-/-} mice also appeared normal (Fig. 3a).

To investigate the molecular mechanism by which the lack of PKC- δ function promotes B-cell proliferation, we examined the production of cytokines that affect B-cell growth. Polymerase chain reaction with reverse transcription of RNA (RT-PCR) analysis revealed that the increase in the abundance of interleukin (IL)-6 messenger RNA induced by cell stimulation was markedly greater in PKC- δ ^{-/-} B cells than in wild-type B cells (Fig. 3c). The amounts of other cytokine mRNAs examined, including those for IL-4 (data not shown) and IL-10 (Fig. 3c), did not differ between the B cells derived from the two types of mice. The amount of IL-6 protein released from PKC- δ ^{-/-} B cells on stimulation was also greatly increased compared with that observed with PKC- δ ^{+/-} B cells (Fig. 3d). Transcription of the *IL-6* gene is regulated predominantly by the transcription factors NF- κ B and NF-IL6 (ref. 23). Electrophoretic mobility-shift assay (EMSA) analysis revealed that, although the DNA-binding activity of NF- κ B did not differ between nuclear extracts prepared from stimulated PKC- δ ^{-/-} and PKC- δ ^{+/-} B cells, that of NF-IL6 was increased in the nuclear extract of PKC- δ ^{-/-} B cells (Fig. 3e). Furthermore, overexpression of PKC- δ in the pre-B-cell line WEHI231 resulted in a reduction in the DNA-binding activity of NF-IL6—expression of a kinase-negative mutant of PKC- δ did not show this effect (Fig. 3f). NF-IL6 is phosphorylated by purified PKC *in vitro*, with the major phosphorylated residue being serine 240, which is located within the DNA-binding domain of NF-IL6 (ref. 24). Phosphorylation of NF-IL6 on this residue resulted in a marked decrease in DNA-binding activity²⁴. Thus PKC- δ probably negatively regulates *IL-6* gene transcription through inhibition of the binding of NF-IL6 to DNA.

To evaluate the function of B cells in PKC- δ ^{-/-} mice, we measured the serum concentrations of immunoglobulin isotypes. The concentrations of IgG1 and IgA in the serum of PKC- δ ^{-/-} mice were significantly greater than those in wild-type mice (Fig. 4a). The titres of other immunoglobulin isotypes were similar in PKC- δ ^{-/-} and wild-type mice. An increase in serum immunoglobulin concentration is often a characteristic of auto-immune disease. Although the serum concentration of antibodies to DNA was not significantly increased in PKC- δ ^{-/-} mice (data not shown), we detected high levels of antibodies to chromatin, most of which were

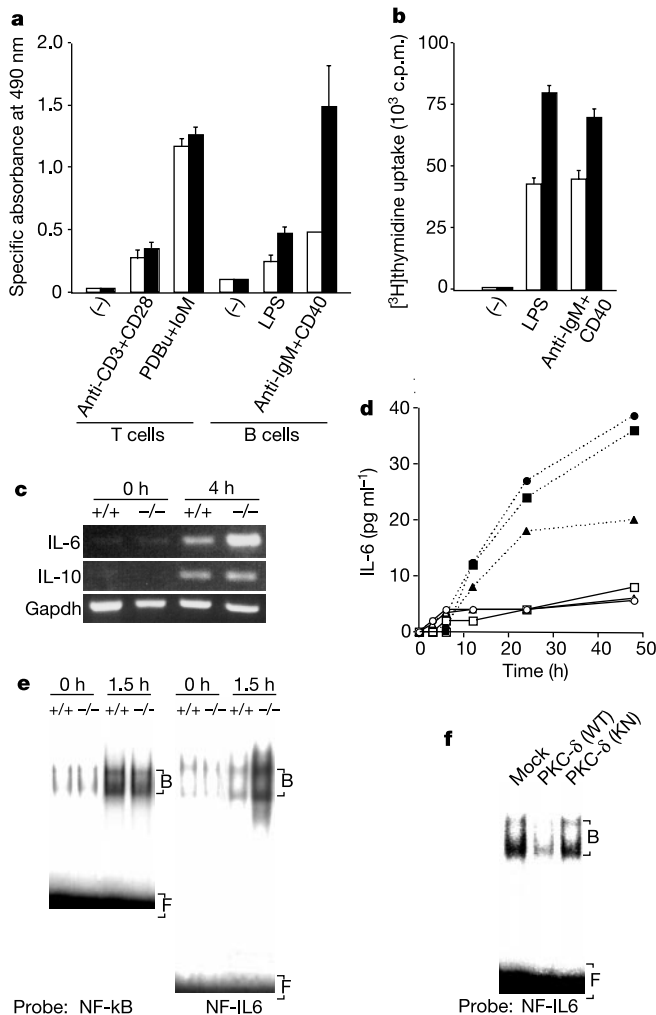


Figure 3 Increased proliferative capacity of B cells from PKC- δ ^{-/-} mice. **a**, Proliferation of splenic T and B cells from wild-type (open bars) and PKC- δ ^{-/-} (filled bars) mice with the indicated stimulation. Cell proliferation was assessed either with an XTT assay after 96 h of stimulation (**a**) or [³H]thymidine incorporation after 48 h of stimulation (**b**). PDBu, phorbol 12,13-dibutyrate; IoM, ionomycin. **c**, d, Increased production of IL-6 by PKC- δ ^{-/-} B cells. Splenic B cells from PKC- δ ^{+/-} (open symbols) or PKC- δ ^{-/-} (filled symbols) mice were cultured in the presence of anti-IgM and anti-CD40 for the indicated times. The amounts of mRNAs were evaluated by RT-PCR (**c**) and the concentration of IL-6 in the culture supernatants was determined by ELISA (**d**). The different symbols indicate different individuals. **e**, **f**, Increase in DNA-binding activity of NF-IL6 induced by the lack of PKC- δ . **e**, Splenic B cells from PKC- δ ^{+/-} or PKC- δ ^{-/-} mice were stimulated with LPS for the indicated times, after which a nuclear extract of the cells was subjected to EMSA with a probe specific for NF- κ B or for NF-IL6. **f**, WEHI231 cells were infected either with recombinant retroviruses coding for wild-type (WT) or kinase-negative (KN) forms of PKC- δ or with the empty vector (Mock). Nuclear extracts of the cells were then subjected to EMSA with the probe for NF-IL6. The positions of free (F) and bound (B) probes are shown.

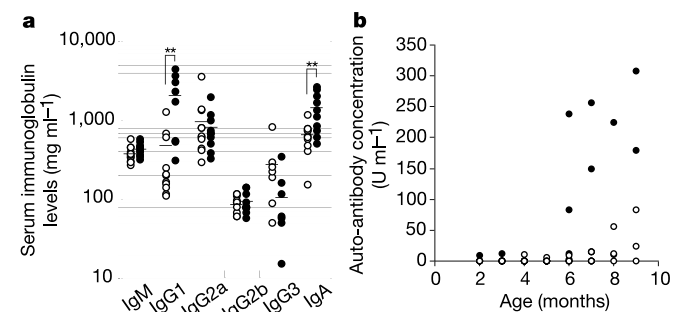


Figure 4 Increased serum concentrations of IgG1, IgA and auto-antibodies in PKC- δ ^{-/-} mice. **a**, Blood was obtained from the tail vein of PKC- δ ^{+/-} (open circles) and PKC- δ ^{-/-} (filled circles) mice, and the serum concentrations of immunoglobulin isotypes were determined by ELISA. Each data point is from an individual mouse and mean values are indicated by short horizontal lines. Double asterisk, $P < 0.01$ (Student's *t*-test). **b**, Blood was obtained from PKC- δ ^{+/-} or PKC- δ ^{-/-} (open circles) and PKC- δ ^{-/-} (filled circles) mice at the indicated age, and the concentration of antibodies to chromatin was determined by ELISA.

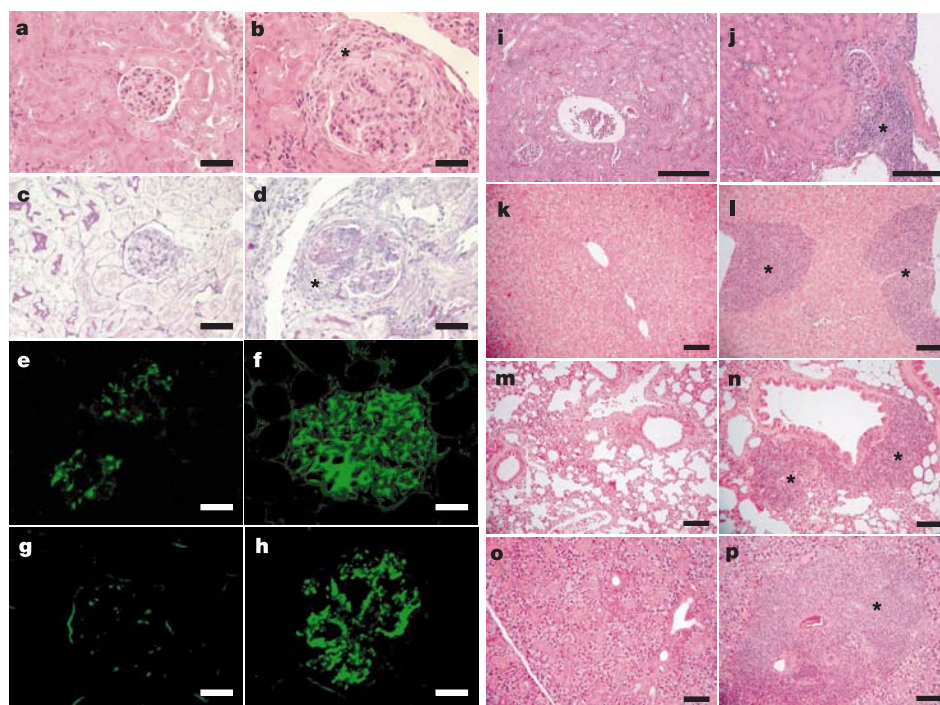


Figure 5 Glomerulonephritis and perivascular infiltration of leukocytes in $PKC-\delta^{-/-}$ mice. **a–h**, Sections of kidney from $PKC-\delta^{+/+}$ (**a, c, e, g**) and $PKC-\delta^{-/-}$ (**b, d, f, h**) mice subjected to staining with haematoxylin and eosin (**a, b**) or periodic acid-Schiff solution (**c, d**), or to immunofluorescence staining with anti-IgG (**e, f**) or anti-C3 (**g, h**). Asterisks

indicate crescent formation. Scale bars, 50 μm . **i–p**, Sections of kidney (**i, j**), liver (**k, l**), lung (**m, n**) and salivary gland (**o, p**) of wild-type (**i, k, m, o**) and $PKC-\delta^{-/-}$ (**j, l, n, p**) mice stained with haematoxylin and eosin. Asterisks indicate abnormal perivascular infiltration of leukocytes. Scale bars: 50 μm (**i, j**), 100 μm (**k–p**).

of the IgG isotype, in the serum of $PKC-\delta^{-/-}$ mice older than 6 months (Fig. 4b).

Histological examination of the kidneys of $PKC-\delta^{-/-}$ mice revealed evidence of glomerulonephritis. The glomeruli were enlarged as a result of mesangial cell proliferation, with occasional crescent formation (Fig. 5a–d). Immunohistochemical staining revealed deposition of IgG and complement component C3 in the mesangial region and along capillary walls of $PKC-\delta^{-/-}$ mice (Fig. 5e–h). Furthermore, perivascular infiltration of leukocytes was detected in the kidney, liver, lungs and salivary gland of $PKC-\delta^{-/-}$ mice (Fig. 5i–p). Immunohistochemical analysis demonstrated that the infiltrating leukocytes comprised mostly B lymphocytes, with a smaller number of T lymphocytes also present (data not shown). These observations suggest that B-cell tolerance is abrogated in $PKC-\delta^{-/-}$ mice as a result of the loss of the anti-proliferative activity of PKC- δ .

Crosslinking of antigen receptors on T and B lymphocytes results in the activation of PKC, suggesting that this enzyme has an important function in transduction of mitogenic signals⁴. Consistent with this hypothesis, mice lacking PKC- β develop an immunodeficiency owing to reduced proliferation of B cells²⁵. Similarly, TCR-mediated T-cell growth was impaired in PKC- θ -deficient mice²⁶. Thus, activation of PKC- β in B cells and of PKC- θ in T cells seems to be necessary for growth stimulation. The phenotype of $PKC-\delta^{-/-}$ mice described here contrasts with that of $PKC-\beta^{-/-}$ mice, suggesting that crosslinking of antigen receptors on B cells results in the activation not only of the pro-mitogenic PKC- β but also of the anti-mitogenic PKC- δ (see Fig. 4 of Supplementary Information), possibly allowing for fine regulation of the immune reaction. Our data suggest that PKC- δ inhibits B-cell growth through transcriptional regulation of the *IL-6* gene. Consistent with these observations, the phenotype of $PKC-\delta^{-/-}$ mice is similar to that of transgenic mice that express the human *IL-6* gene; the

latter mice thus exhibit B-cell expansion, an increased serum concentration of IgG1, mesangio-proliferative glomerulonephritis, and infiltration of B cells in many organs²⁷. Given that the lack of PKC- δ leads to hyperproliferation of B cells and auto-immunity in mice, it is possible that the activity of PKC- δ is reduced in the B cells of humans with certain auto-immune diseases, such as Castleman's disease, and agents that modulate PKC- δ activity may therefore be of therapeutic value in such individuals. □

Methods

Generation of PKC- δ -deficient mice

We isolated the genomic DNA of the *PKC- δ* gene from a 129/Sv mouse genomic library (Stratagene). The targeting vector was constructed by replacing a 2.8-kilobase (kb) *HpaI*-*Bam*HI fragment containing exons I and II of the *PKC- δ* gene with a PGK-neo-poly(A) cassette. The targeting vector thus contained 1.2- and 6.2-kb regions of homology 5' and 3' of the neomycin-resistance marker, respectively. The PGK-TK-poly(A) cassette was ligated at the 3' end of the insert. The maintenance, transfection and selection of embryonic stem cells were performed as described²⁸. The mutant embryonic stem cells were microinjected into C57BL/6 blastocysts, and the resulting male chimaeras were mated with female C57BL/6 mice. Heterozygous offspring were intercrossed to produce homozygous mutant animals. All mice were maintained in a specific pathogen-free animal facility.

Immunoblot analysis

We performed an immunoblot analysis as described^{28,29}. The blots were then probed with mouse monoclonal antibodies to mouse PKC- δ , to heat shock protein of 90K (Hsp90), or to glycogen synthase kinase-3 β (GSK-3 β) (Transduction Laboratories). Immune complexes were detected with the ECL blotting system (Amersham Pharmacia Biotech).

Histopathology and immunofluorescence staining

Organs were removed from necropsied animals, fixed in 4% paraformaldehyde, embedded in paraffin, and stained with haematoxylin and eosin or periodic acid-Schiff solution. For three-colour immunofluorescence staining, frozen sections were incubated for 30 min at room temperature first with biotinylated rat monoclonal antibodies to CD4 and to CD8, and then with Cy5.5-conjugated streptavidin (Cortex Biochemicals). After washing with phosphate-buffered saline (PBS), the sections were stained for 30 min with a rat Alexa488-labelled monoclonal antibody to B220 and tetramethylrhodamine-conjugated PNA

(Vector Laboratories). To detect deposition of immune complexes at glomeruli, we incubated frozen sections for 30 min at room temperature with fluorescein isothiocyanate (FITC)-labelled goat antibodies to mouse IgG or to C3 (ICN Pharmaceuticals). We performed immunohistochemistry as described²⁹.

Flow cytometry

Single-cell suspensions were prepared from spleen, lymph nodes, bone marrow and thymus, and were stained with a combination of FITC-conjugated antibodies to B220 and phycoerythrin-conjugated antibodies to TCR β , with FITC-conjugated antibodies to B220 and phycoerythrin-conjugated antibodies to IgM, or with FITC-conjugated antibodies to CD8 and cychrome-conjugated antibodies to CD4. All antibodies used for flow cytometry were obtained from PharMingen. We performed all analyses with a FACSCalibur flow cytometer and Cell Quest software (Becton Dickinson).

Lymphocyte purification and proliferation

For isolation and purification of B lymphocytes, single-cell suspensions prepared from spleen or lymph nodes were treated for 30 min at 4 °C with antibodies to mouse Thy-1.2 (Serotec), and then subjected to lysis for 1 h at 37 °C with rabbit low-toxicity complement (Wako). After removal of dead cells by centrifugation in lympholyte-M (Cedarlane), the remaining cells were plated on plastic dishes for 1 h at 37 °C to remove adherent macrophages. The purity of the resulting B-cell preparations was usually greater than 90% and did not differ substantially between PKC- $\delta^{+/+}$ and PKC- $\delta^{-/-}$ mice (see Fig. 5 of Supplementary Information). Purification and analysis of the proliferation of T cells were performed as described²⁹. Purified B cells (5×10^5) were cultured in the wells of 96-well plates containing complete RPMI supplemented either with goat F(ab')₂ specific for mouse IgM (Jackson Immuno Research) and a hamster monoclonal antibody to mouse CD40 (PharMingen), each at a final concentration of 10 μ g ml⁻¹, or with LPS (Sigma) at a final concentration of 0.2 μ g ml⁻¹. We assessed cell proliferation after 96 h of culture with the use of an XTT colorimetric assay^{28,29} (Boehringer Mannheim). The rate of B-cell proliferation was also evaluated by measurement of the incorporation of [³H]thymidine. Purified B cells were cultured as described above for 48 h, the final 3 h of the incubation being performed in the presence of 1 μ Ci [³H]thymidine per well. Radioactivity corresponding to the incorporated [³H]thymidine was measured with a 1205 BETAPLATE (Wallac).

Adoptive transfer of B cells

PKC- $\delta^{+/+}$ and PKC- $\delta^{-/-}$ mice at 9 weeks of age were used as B-cell donors. B cells were isolated and purified as described above. The transferred B cells from the wild-type and mutant mice appeared indistinguishable—they were IgM⁺ IgD⁺ B220^{hi} CD5⁻ mature re-circulating B cells. Furthermore, the expression of CD86 (B7-2), an activation marker of B cells, did not differ between the two cell populations (either in terms of the percentage of cells expressing this marker or with regard to the level of expression). The purified B cells (1.5×10^7) in 100 μ l PBS were administered intravenously into the orbital venous plexus of recipient Rag1^{-/-} mice (Jackson Laboratory) at 8 weeks of age. The recipient mice were killed and analysed 18 days after B-cell transfer.

Quantification of serum immunoglobulins and cytokines

The amounts of each immunoglobulin isotype in serum were determined by enzyme-linked immunosorbent assay (ELISA) with antibodies specific for each isotype (monoclonal antibody-based mouse immunoglobulin isotyping kit; PharMingen). Auto-antibodies to chromatin were measured by ELISA as described³⁰. Auto-antibody concentration was expressed in units, with reference to a standard curve obtained with a standard serum pool from (New Zealand Black $\times$ New Zealand White) F₁ mice. For analysis of cytokine mRNAs or protein, purified B cells (1×10^6) were cultured for the indicated times in the wells of 24-well plates containing complete RPMI supplemented with goat F(ab')₂ specific for mouse IgM and a hamster mAb to mouse CD40, each at a final concentration of 10 μ g ml⁻¹. Transcripts of IL-6, IL-4, IL-10 and glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*) genes were detected with an RT-PCR kit (ReverTra Ace, Toyobo). For measurement of IL-6 protein released from the B cells, the culture supernatants were subjected to ELISA with antibodies specific for IL-6 (Biosource International).

EMSA analysis

We stimulated B cells with LPS for 90 min as described above. WEHI231 cells were infected either with recombinant retroviruses coding for wild-type or kinase-negative PKC- δ^5 , or with the empty retroviral vector alone. Nuclear extracts prepared from the stimulated B cells or from the infected WEHI231 cells were pre-incubated for 10 min at room temperature in a final volume of 20 μ l containing 10 μ g sonicated salmon sperm DNA (Sigma), 10% glycerol, 25 mM HEPES-KOH (pH 7.9), 0.5 mM EDTA, 0.5 mM dithiothreitol, and 50 mM KCl. A ³²P-labelled oligonucleotide probe (50,000 c.p.m.) containing binding sites for NF- κ B or NF-IL6 (refs. 24,26) was added to the reaction mixture, which was then incubated for 30 min at 30 °C. The formation of protein-probe complexes was analysed by electrophoresis through a 5% polyacrylamide gel. The gel was then dried and subjected to autoradiography with an intensifying screen.

Received 26 October 2001; accepted 4 March 2002.

1. Nishizuka, Y. Protein kinase C and lipid signaling for sustained cellular responses. *FASEB J.* **9**, 484–496 (1995).
2. Mellor, H. & Parker, P. J. The extended protein kinase C superfamily. *Biochem. J.* **332**, 281–292 (1998).
3. Livneh, E. & Fishman, D. D. Linking protein kinase C to cell-cycle control. *Eur. J. Biochem.*

248, 1–9 (1997).

4. Weiss, A. & Littman, D. R. Signal transduction by lymphocyte antigen receptors. *Cell* **76**, 263–274 (1994).
5. Ohno, S., Akita, Y., Konno, Y., Imajoh, S. & Suzuki, K. A novel phorbol ester receptor/protein kinase, nPKC, distantly related to the protein kinase C family. *Cell* **53**, 731–741 (1988).
6. Gschwendt, M. Protein kinase C δ . *Eur. J. Biochem.* **259**, 555–564 (1999).
7. Watanabe, T. *et al.* Cell division arrest induced by phorbol ester in CHO cells overexpressing protein kinase C- δ subspecies. *Proc. Natl Acad. Sci. USA* **89**, 10159–10163 (1992).
8. Mischak, H. *et al.* Overexpression of protein kinase C- δ and - ϵ in NIH 3T3 cells induces opposite effects on growth, morphology, anchorage dependence, and tumorigenicity. *J. Biol. Chem.* **268**, 6090–6096 (1993).
9. Fukumoto, S. *et al.* Protein kinase C δ inhibits the proliferation of vascular smooth muscle cells by suppressing G1 cyclin expression. *J. Biol. Chem.* **272**, 13816–13822 (1997).
10. Mishima, K., Ohno, S., Shitara, N., Yamaoka, K. & Suzuki, K. Opposite effects of the overexpression of protein kinase C γ and δ on the growth properties of human glioma cell line U251 MG. *Biochem. Biophys. Res. Commun.* **201**, 363–372 (1994).
11. Harrington, E. O. *et al.* Enhancement of migration by protein kinase C α and inhibition of proliferation and cell cycle progression by protein kinase C δ in capillary endothelial cells. *J. Biol. Chem.* **272**, 7390–7397 (1997).
12. Pessino, A. *et al.* Antisense oligodeoxynucleotide inhibition of δ protein kinase C expression accelerates induced differentiation of murine erythroleukaemia cells. *Biochem. J.* **312**, 549–554 (1995).
13. Li, W. & Pierce, J. H. Protein kinase C- δ , an important signaling molecule in the platelet-derived growth factor β receptor pathway. *Curr. Topics Microbiol. Immunol.* **211**, 55–65 (1996).
14. Mischak, H. *et al.* Phorbol ester-induced myeloid differentiation is mediated by protein kinase C- α and - δ and not by protein kinase C- β II, - ϵ , - ζ , and - η . *J. Biol. Chem.* **268**, 20110–20115 (1993).
15. Emoto, Y., Kikaki, H., Manome, Y., Kharbanda, S. & Kufe, D. Activation of protein kinase C δ in human myeloid leukemia cells treated with 1- β -D-arabinofuranosylcytosine. *Blood* **87**, 1990–1996 (1996).
16. Emoto, Y. *et al.* Proteolytic activation of protein kinase C δ by an ICE-like protease in apoptotic cells. *EMBO J.* **14**, 6148–6156 (1995).
17. Ghayur, T. *et al.* Proteolytic activation of protein kinase C δ by an ICE/CED 3-like protease induces characteristics of apoptosis. *J. Exp. Med.* **184**, 2399–2404 (1996).
18. Li, W. *et al.* Expression of an ATP binding mutant of PKC- δ inhibits Sis-induced transformation of NIH3T3 cells. *Oncogene* **13**, 731–737 (1996).
19. La Porta, C. A. & Comolli, R. Over-expression of protein kinase C δ is associated with a delay in preneoplastic lesion development in diethylnitrosamine-induced rat hepatocarcinogenesis. *Carcinogenesis* **16**, 1233–1238 (1995).
20. Geiges, D., Marks, F. & Gschwendt, M. Loss of protein kinase C δ from human HaCaT keratinocytes upon ras transfection is mediated by TGF α . *Exp. Cell Res.* **219**, 299–303 (1995).
21. Lu, Z. *et al.* Tumor promotion by depleting cells of protein kinase C δ . *Mol. Cell. Biol.* **17**, 3418–3428 (1997).
22. Reddig, P. J. *et al.* Transgenic mice overexpressing protein kinase C δ in the epidermis are resistant to skin tumor promotion by 12-O-tetradecanoylphorbol-13-acetate. *Cancer Res.* **59**, 5710–5718 (1999).
23. Akira, S. & Kishimoto, T. IL-6 and NF- κ B in acute-phase response and viral infection. *Immunol. Rev.* **127**, 25–50 (1992).
24. Trautwein, C., van der Geer, P., Karin, M., Hunter, T. & Chojkier, M. Protein kinase A and C site-specific phosphorylations of LAP (NF-IL6) modulate its binding affinity to DNA recognition elements. *J. Clin. Invest.* **93**, 2554–2561 (1994).
25. Leitges, M. *et al.* Immunodeficiency in protein kinase C β -deficient mice. *Science* **273**, 788–791 (1996).
26. Sun, Z. *et al.* PKC- θ is required for TCR-induced NF- κ B activation in mature but not immature T lymphocytes. *Nature* **404**, 402–407 (2000).
27. Suematsu, S. *et al.* IgG1 plasmacytosis in interleukin 6 transgenic mice. *Proc. Natl Acad. Sci. USA* **86**, 7547–7551 (1989).
28. Nakayama, K. *et al.* Mice lacking p27^{Kip1} display increased body size, multiple organ hyperplasia, retinal dysplasia, and pituitary tumors. *Cell* **85**, 707–720 (1996).
29. Tsukiyama, T. *et al.* Down-regulation of p27^{Kip1} expression is required for development and function of T cells. *J. Immunol.* **166**, 304–312 (2001).
30. Burlingame, R. W. & Rubin, R. L. Subnucleosome structures as substrates in enzyme-linked immunosorbent assays. *J. Immunol. Methods* **134**, 187–199 (1990).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank S. Akira and K. Hoshino for the plasmids; N. Motoyama, A. Yoshimura and U. Kikkawa for discussions; Y. Yamada, K. Shimoharada and N. Ohtsui for technical assistance; and M. Kimura and C. Sugita for help in preparation of the manuscript. This work was supported in part by a grant from the Ministry of Education, Science, Sports and Culture of Japan, by Nissan Science Foundation, and by a research grant from the Human Frontier Science Program.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.I.N. (e-mail: nakayak1@bioreg.kyushu-u.ac.jp).

AP-1 functions upstream of CREB to control synaptic plasticity in *Drosophila*

Subhabrata Sanyal^{*†}, David J. Sandstrom^{*†‡}, Charles A. Hoeffler^{*} & Mani Ramaswami^{*}

^{*} Department of Molecular and Cellular Biology, and Arizona Research Laboratories, Division of Neurobiology, University of Arizona, Tucson, Arizona 85721, USA

[†] These authors contributed equally to this work

[‡] Present address: Laboratory of Molecular Biology, NIMH Building 36, Room 1B10, National Institutes of Health, Bethesda, Maryland 20892, USA

Activity-regulated gene expression mediates many aspects of neural plasticity, including long-term memory. In the prevailing view, patterned synaptic activity causes kinase-mediated activation of the transcription factor cyclic AMP response-element-binding protein, CREB. Together with appropriate cofactors,

CREB then transcriptionally induces a group of 'immediate-early' transcription factors and, eventually, effector proteins that establish or consolidate synaptic change¹. Here, using a *Drosophila* model synapse, we analyse cellular functions and regulation of the best known immediate-early transcription factor, AP-1; a heterodimer of the basic leucine zipper proteins Fos and Jun². We observe that AP-1 positively regulates both synaptic strength and synapse number, thus showing a greater range of influence than CREB³. Observations from genetic epistasis and RNA quantification experiments indicate that AP-1 acts upstream of CREB, regulates levels of CREB messenger RNA, and functions at the top of the hierarchy of transcription factors known to regulate long-term plasticity. A Jun-kinase signalling module provides a CREB-independent route for neuronal AP-1 activation; thus, CREB regulation of AP-1 expression⁴ may, in some neurons, constitute a positive feedback loop rather than the primary step in AP-1 activation.

Synaptic signalling modulates a transcriptional cascade gated by the cyclic AMP-responsive element-binding protein (CREB) and immediate-early transcription factors such as AP-1, c/EBP and Zif-268 to control gene expression that underlies persistent forms of synaptic plasticity¹. A role for CREB at the top of the transcriptional

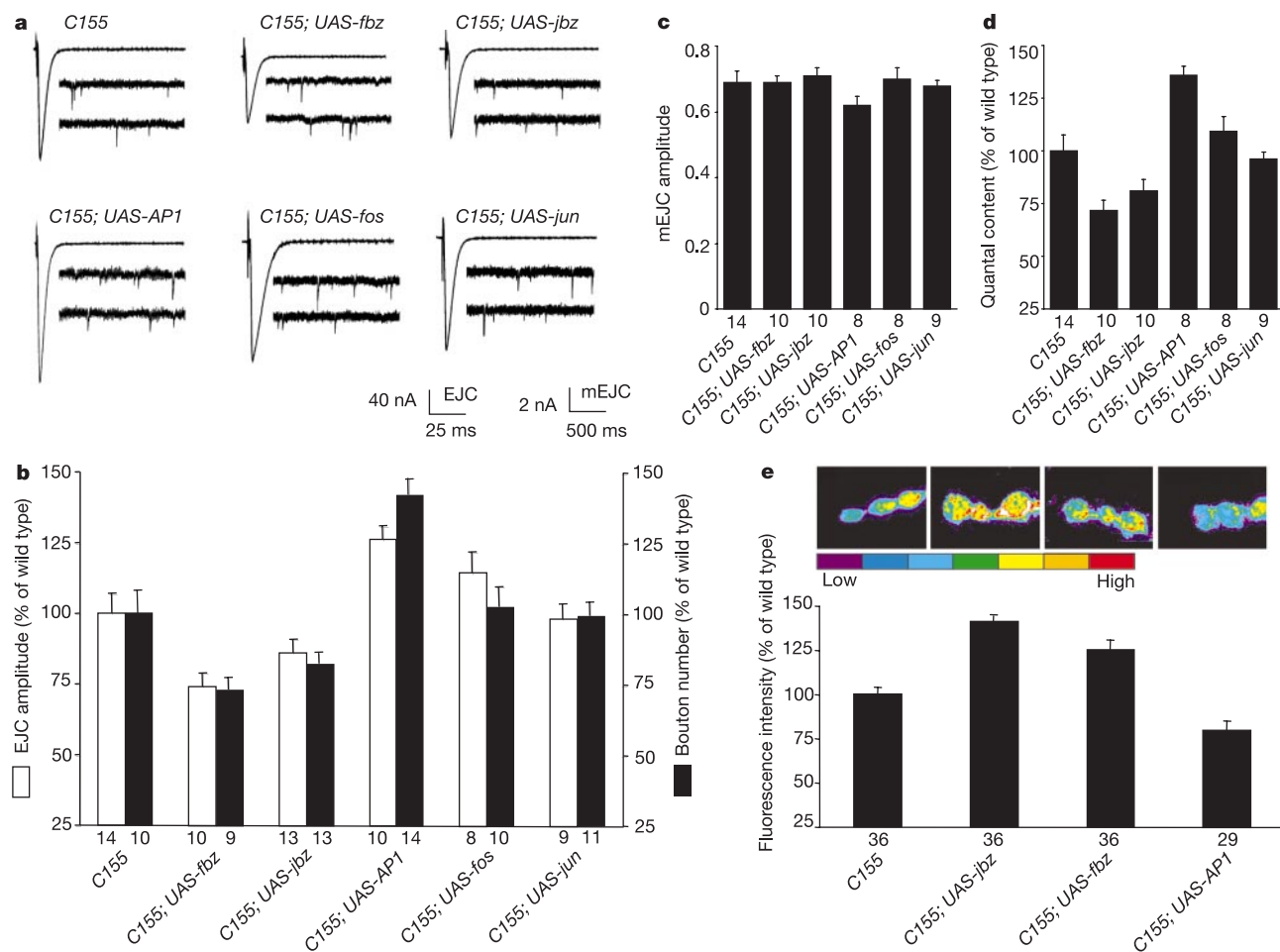


Figure 1 Neuronal AP-1 controls bouton number and synaptic strength. **a**, **b**, EJCs and mEJC traces (**a**), and histogram representation of EJC amplitude and bouton number (**b**) from larvae in which either wild-type or dominant inhibitory forms of Fos (Fbz) or Jun (Jbz) are expressed. Fbz causes a 30% reduction and Jbz a 25% reduction in synapse size ($P < 0.005$ and $P < 0.02$, respectively) and strength ($P < 0.001$ and $P < 0.03$, respectively). Although Fos or Jun alone have minimal or no effects on bouton number and synaptic strength, together they cause an increase of 30% ($P < 0.001$ for bouton number and synaptic strength). **c**, **d**, Miniature (m)EJC sizes in all genetic combinations tested are

not significantly altered (**c**); thus, alterations in synapse strength arise from reduced quantal content (**d**). **e**, The pseudocolour scale indicates levels of synaptic fasciclin II (top panel). The histogram shows quantitative fluorescence measurements indicating increased fasciclin II in Fbz- ($P < 0.001$) and Jbz- ($P < 10^{-5}$) expressing preparations, and decreased fasciclin II in AP-1-induced preparations ($P < 0.004$). The number of larvae examined is indicated below the bars. For **e**, the number of type 1b termini (4–6 per larva) is indicated.

cascade is strongly indicated by demonstrations that induction of an activator isoform of CREB substantially enhances long-term memory in *Drosophila* and rodents, and long-term facilitation in the sea slug *Aplysia*^{5–7}. These findings, together with substantial data to indicate the CREB responsiveness of several immediate-early genes⁸, has led to a 'switch' model for CREB, in which CREB acts upstream of all transcriptional processes required to effect enduring synaptic change¹. Motivated by observations potentially inconsistent with this model^{3,9}, we investigated neural functions of AP-1—a heterodimeric transcription factor composed of Fos and Jun—and contrasted these to known functions of CREB. Detailed analysis of a Fos-, Jun- and Jun N-terminal kinase (JNK) signalling module's function in *Drosophila* development provides a genetic framework¹⁰ for a dissection of AP-1 function during synaptic plasticity at the *Drosophila* larval neuromuscular junction (NMJ). The maturation of this synapse as it grows in size (number of boutons) and strength (stimulus evoked response) during the three instars of larval growth

involves activity-dependent mechanisms that also underlie long-term plasticity and memory¹¹. Thus, genetic manipulation of neuronal activity or cAMP levels results in predictable alterations in synaptic size and strength¹². Hyperactivity- or cAMP-induced increases in synaptic strength are CREB dependent; this requires increases in bouton number that occur through a reduction in levels of fasciclin II, a synapse cell adhesion molecule whose orthologue in *Aplysia* is rapidly downregulated after induction of long-term facilitation^{3,13,14}.

Our results show that neuronal AP-1 positively regulates both bouton number and synaptic strength; such extensive control of plasticity processes is unique among known transcription factors. Increased neural expression of both Fos and Jun, but not of either alone, results in a 30% potentiation in evoked junctional currents (EJCs) and a parallel increase in the number of synaptic boutons (Fig. 1a, b). Inhibition of either protein by means of targeted neural expression of Fbz and Jbz, dominant negative forms of Fos and Jun,

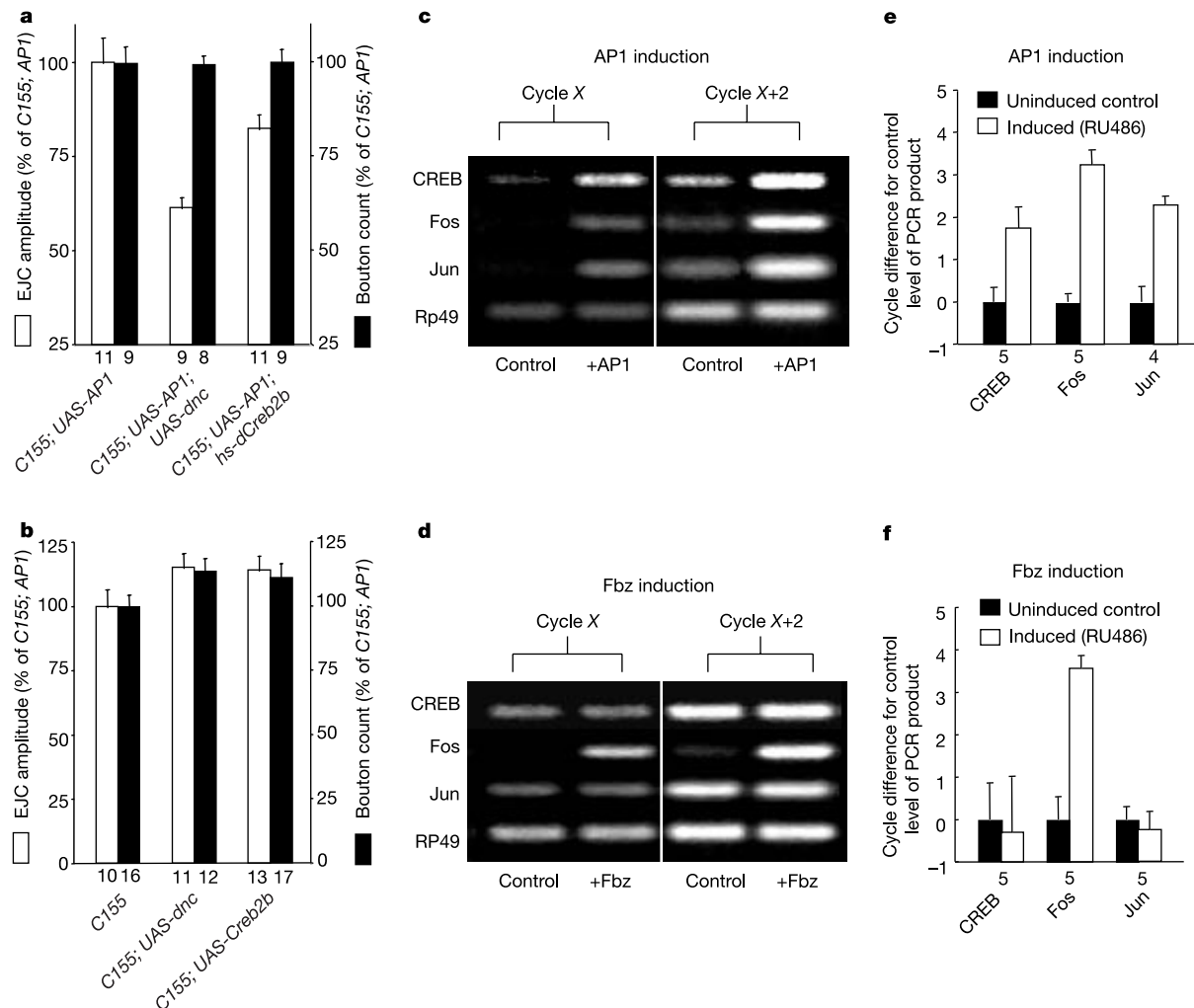


Figure 2 Both CREB and a cAMP-sensitive activity are required downstream of AP-1 to regulate synapse strength but not bouton number. **a**, In a genetic background where AP-1 is induced in neurons, induction of either an inhibitory isoform of CREB (through *hs-Creb2b*) or cAMP phosphodiesterase (through *UAS-dnc*) does not perturb the effect of AP-1 on bouton number. In contrast, both *hs-Creb2b* ($P < 0.02$) and *UAS-dnc* ($P < 0.001$) significantly inhibit the effect of AP-1 on synaptic strength. (*C155; AP-1; hs-dCreb2b* animals have 108%, and *C155; AP-1; UAS-dnc* have 83% average EJC amplitude as compared with wild type.) **b**, In a wild-type background, identical induction of *hs-Creb2b* or neuronal expression of *UAS-dnc* has no effect on either synapse size or

strength. **c–f**, AP-1 induction in neurons leads to rapid threefold upregulation of CREB2 mRNA. **c**, CREB2 transcript levels are elevated in AP-1-induced neurons when compared with uninduced control animals. **d**, Ru486 induction of Fbz does not significantly alter levels of Jun and CREB2 mRNA. **e**, **f**, Quantitative comparison of mRNA levels by Q-PCR. After induction of AP-1, *Cerb2* mRNA levels increase threefold relative to uninduced controls ($P < 0.02$). **f**, After Fbz induction, Fbz transcript levels are increased 12-fold whereas the average levels of Jun and CREB2 remain unchanged when compared to mRNA from uninduced controls.

respectively¹⁵, causes opposite phenotypes at the larval NMJ (a roughly 30% reduction in bouton number and synaptic strength, Fig. 1a, b). The similarity and specificity of these effects indicate that Fbz and Jbz act by specifically inhibiting endogenous AP-1 activity, a premise supported by demonstrations that developmental phenotypes caused by induction of Fbz or Jbz are identical to those caused by loss-of-function mutations in D-Fos or D-Jun¹⁶. The observation that Fos and Jun positively regulate synapse plasticity in a manner that requires the presence of both molecules, but is sensitive to inhibition of either, indicates that Fos and Jun function as a heterodimer to control synapse plasticity. This is important, given various data to suggest that AP-1 is a heterogeneous collection of molecularly distinct transcription factors¹⁷. Expression of AP-1, Fbz or Jbz through a promoter relatively restricted to motor neurons¹⁸ was sufficient for the observed effects; perturbation of AP-1 activity in postsynaptic muscle had little or no effect on the synaptic parameters analysed (data not shown). Together these data are consistent with a cell-autonomous role for AP-1 in regulating neural plasticity.

AP-1 regulates synapse strength by controlling the quantal content (number of vesicles fusing per stimulus) of presynaptic transmitter release (Fig. 1c). Thus, whereas AP-1 manipulations affect EJCs, responses to individual synaptic vesicles (miniature (m)EJC amplitudes) remain unchanged (Fig. 1a, c). Synapse expansion mediated by AP-1 is accompanied by reduced (by approximately 25%) levels of synaptic fasciclin II; Fbz- or Jbz-mediated reductions in synapse size are accompanied by increased (by about 35%) levels of synaptic fasciclin II (Fig. 1e). This is consistent with a model in which AP-1-mediated changes in synapses occur through the classical activity-regulated pathway for synapse plasticity¹⁴.

Previous work has demonstrated a requirement for CREB in determining synaptic strength but not bouton number at the *Drosophila* larval NMJ³. In this study, the observed consequences of AP-1 manipulations were not consistent with a simple model of CREB exerting its effects through Fos and Jun expression. Rather, they suggested a model in which CREB acts downstream of AP-1 in the pathway, leading to altered synaptic strength. We tested this hypothesis by asking whether induction of a CREB-blocker¹⁹ transgene (*hs-Creb2b*) would inhibit AP-1-induced synaptic changes. Induction of *hs-Creb2b* (ref. 3) had no effect on synaptic size and strength in a control genetic background; however, it specifically blocked the AP-1-induced increase of synaptic strength, without affecting the AP-1-induced increase of bouton number (Fig. 2). This specific effect of CREB blocker strongly suggests that the effect of AP-1 on synaptic strength occurs through a pathway that involves the activation of CREB; however, its effect on bouton number occurs by means of the CREB-independent pathway, suggested by previous studies.

One formal explanation for these observations is that AP-1 induction in neurons results in general hyperexcitability; that is, it acts to turn on the activity- and cAMP-dependent signalling pathway in neurons. In this case, reduction of neuronal cAMP through targeted overexpression of the *dunce* (*dnc*) cAMP phosphodiesterase¹⁸ would be expected to block all effects of AP-1 overexpression. This was not the case. Neural induction of cAMP phosphodiesterase did not alter the influence of AP-1 on bouton number (Fig. 2b); this is consistent with a function of AP-1 downstream of the cAMP-regulated stage in structural synaptic plasticity. In contrast, reduced neural cAMP completely blocked the effect of AP-1 on synaptic strength. As with the CREB-blocker transgene, identical induction of cAMP phosphodiesterase had no effect on synaptic size or strength in a control genetic background; thus, low cAMP specifically affects the AP-1-induced potentiation of synaptic strength. The simplest interpretation of these observations is that the activity of AP-1 is not substantially reduced at low levels of cAMP achieved in our experiment; however, activity of a downstream molecule required for the effect of AP-1 on synaptic strength is inhibited

under these conditions. These observed consequences of cAMP and CREB inhibition on the effects of AP-1 confirm and extend previous studies demonstrating independent control of synapse size and strength³. The similarity of low cAMP and CREB-inhibition effects on AP-1-induced plasticity (Fig. 2) is consistent with a model in which the downstream target of cAMP signalling is CREB itself.

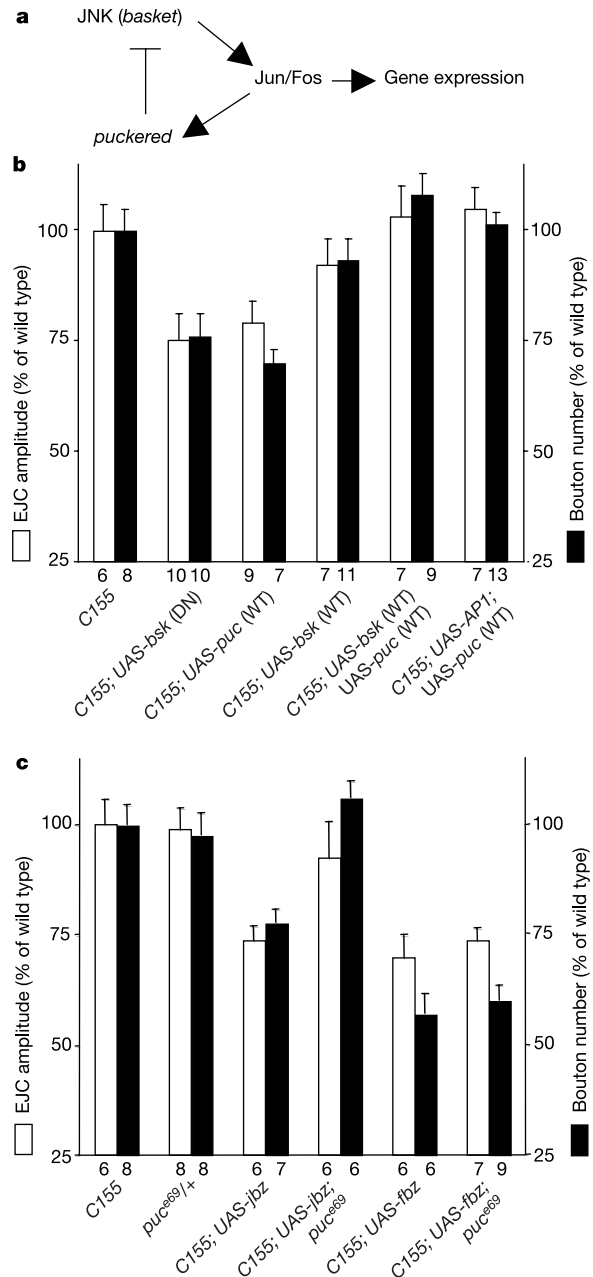


Figure 3 JNK signalling regulates AP-1-dependent synaptic plasticity. **a**, Outline of proximal components of the JNK signalling module in *Drosophila*¹⁰. **b**, The JNK signalling pathway positively regulates both structural and functional synaptic plasticity. JNK inhibition through neural expression of either a dominant negative JNK (*C155; UAS-bsk* (DN)) or wild-type JNK-phosphatase (*C155; UAS-puc* (WT)) reduces synapse size ($P < 0.0001$ and $P < 0.0001$) and strength ($P < 0.04$ and $P < 0.03$) by about 25–30%. These effects of Puc are suppressed by simultaneous overexpression of wild-type JNK. Neural overexpression of Puc neutralizes all effects of AP-1 induction ($P < 0.01$). **c**, Genetic reduction of JNK phosphatase in flies heterozygous for a *puc* hypomorphic allele (*puc^{eb9}/+*) has no observable effect in a wild-type background, but restores normal bouton number and EJC amplitude to Jun-inhibited motor neurons.

One potential mechanism for AP-1 regulation of CREB is suggested by the presence of consensus AP-1-binding sites in the *Drosophila* CREB2 promoter. Consistent with a direct transcriptional control mechanism, we observed a robust increase in CREB mRNA levels after a brief pulse of increased neural AP-1 expression. Brief induction of AP-1 in the *Drosophila* nervous system was achieved using the *elavGS-GAL4* transgene, which causes pan-neuronal expression of a steroid-dependent, inactive variant of the Gal4 transcription factor. In *elavGS-GAL4; UAS-AP-1* animals, a roughly ninefold increase in levels of neural AP-1 is induced by a 6-h incubation on food doped with the steroid analogue Ru486 (Fig. 2c)²⁰. Under these conditions, CREB mRNA levels are increased about threefold. That this increase is caused by AP-1, rather than unidentified consequences of doping with Ru486 for instance, is evidenced by the unchanged levels of CREB mRNA in Ru486-fed animals expressing Fbz under *elavGS-GAL4* control (Fig. 2d).

Our proposal that AP-1 functions upstream of CREB to regulate synaptic plasticity, predicts the existence of a CREB-independent signalling pathway that can act as a primary activator of neuronal AP-1. One candidate is the JNK pathway, known to regulate AP-1 function in several contexts in mammals and *Drosophila*^{21,22}. Several reports^{23–25} point to the existence of a Ca²⁺-dependent route to JNK activation in the nervous system. Two critical elements of the JNK pathway expressed in the *Drosophila* nervous system are *basket* (*bsk*), which codes for JNK²², and *puckered* (*puc*), which codes for a phosphatase that, by dephosphorylating JNK, negatively regulates JNK signalling²⁶ (Fig. 3a). We tested how altered JNK signalling affected synapse plasticity. Reduced JNK signalling by neural overexpression of either Puckered or dominant negative Basket (*Bsk*^{DN}), mimicked the synaptic effects of AP-1 inhibition. Thus both treatments result in a roughly 30% decrease in synapse strength and bouton number (Fig. 3b). Although overexpression of Basket JNK causes no phenotype on its own, it completely neutralizes the effect of Puckered overexpression; this indicates that effects of Puckered on synapse plasticity occur through its documented inhibition of JNK signalling (Fig. 3b). In contrast to reduced cAMP (Fig. 2), reduced JNK signalling (achieved by neural expression of Puckered) blocked effects of AP-1 on both structural and functional synaptic plasticity (Fig. 3b). This is consistent with a model in which JNK signals through AP-1 to regulate synaptic change.

To ascertain further that the effects of JNK on synaptic change are mediated by AP-1, we studied the effects of increased JNK signalling on synapses inhibited for AP-1 (Fig. 3c). If JNK acts by increasing Jun function, then increased JNK signalling should alleviate the consequences of AP-1 inhibition that derive from reduced Jun activity. We increased JNK signalling by using a genetic background

heterozygous for *puc*^{E69}, a loss-of-function allele of *puc*. Although the resulting enhancement of JNK signalling was not sufficient to drive synaptic change in a wild-type background, it completely suppressed effects of neural Jbz expression (Fig. 3c). This observation, that reduced endogenous *puc* function can compensate when neuronal AP-1 is inhibited, demonstrates that the neural AP-1 activity *in vivo* is positively regulated by JNK signalling.

We demonstrate that AP-1, under JNK regulation, functions upstream of CREB to control a wider range of plasticity processes than anticipated; indeed, our data are consistent with a model in which AP-1 activation is sufficient for transcriptional control of long-term plasticity. Our finding that AP-1 activates CREB, probably through the regulation of CREB mRNA levels, contrasts with previous demonstrations that CREB positively regulates AP-1 transcription⁴. These apparently conflicting observations may be rationalized if CREB induction of AP-1 is considered as part of a positive feedback loop rather than the primary mechanism for AP-1 activation (Fig. 4). Observations presented here demonstrate unanticipated functions of AP-1 and expose limitations of current models for transcriptional control of long-term plasticity. Further studies in other neuronal subtypes are required to establish the generality of our observations. □

Methods

Drosophila strains and genetic controls

We used the following strains: wild type (Oregon R; D. Brower); Gal4-responsive *UAS-fos*, *UAS-jun*, *UAS-jbz* and *UAS-jbz* transgenes (M. Bienz); *hs-Creb2b* (J. Yin and T. Tully); neural Gal4 line, *C155* and muscle line *MHC-GAL4* (C. Goodman); muscle and neural Gal4 lines, respectively *24B* and *elav-GAL4* (L. Luo); *D42* (G. Boulianne); *UAS-bsk* and *UAS-bsk*^{DN} (M. Mlodzik); *UAS-puc* and *puc*^{E69} (A. Martinez Arias). *ElavGS-GAL4* lines were from T. Osterwalder and H. Keshishian. In general, experimental animals were generated by crossing males homozygous for UAS transgenes with virgin females homozygous for the Gal4 driver; wild-type controls consisted of the Gal4 driver crossed to the wild type (Oregon R). In cases where this was not feasible, larvae of the same sex as the experimental animals, from closely matched genetic backgrounds, served as controls. We induced the CREB-blocker transgene, *hs-Creb2b*, as described previously³. For inducing Gal4 activity in adult *Drosophila* expressing the *elavGS-GAL4* transgene, 1–3-day-old flies were starved for 6 h and transferred to glass bottles with standard food containing 0.02 mg ml⁻¹ Ru486 (Sigma) for 6 h before decapitation and RNA extraction. We reared flies at 25 °C.

Immunocytochemistry and antibodies

Bouton number in segment A2 was assessed by counting DSYT2 (from H. Bellen) immunoreactive varicosities on muscles 6 and 7. Muscle surface areas were not significantly different in all genotypes analysed. To quantify levels of fasciclin II, synapses were fluorescently labelled with 1D4 antibodies (from C. Goodman), and three terminal type 1b boutons of segment A2 or A3 were imaged at maximum resolution (68 nm pixel diameter) using a cooled charge-coupled device camera (Princeton Instruments) and Metamorph imaging software (Universal Imaging). After background subtraction, the average pixel intensity in these boutons and surrounding subsynaptic reticulum was measured and analysed.

Quantitative PCR (Q-PCR)

For Q-PCR, approximately 50 fly heads were collected for each sample; total RNA was extracted using the RNeasy kit (Qiagen) and purified from genomic DNA using the DNA-free DNase kit (Ambion). Two micrograms of total RNA were used to synthesize oligo(dT)-primed complementary DNA with the Omniscript cDNA synthesis kit (Qiagen). The cDNA was diluted 1:5 with nuclease-free H₂O (Invitrogen) for Q-PCR reactions performed on the Cepheid SMARTCycler using reaction ingredients and the standard protocol from the Quantitect kit (Qiagen). Messenger RNAs in all samples were diluted to ensure identical levels of Rp49. Transcript levels were determined using primer sets (details available on request) specific to Rp49, D-fos, D-jun and *Drosophila* (d)CREB2. Each Q-PCR reaction was repeated three times for five independent RNA preparations. For gel visualization, experimental and control PCR with reverse transcription (RT-PCR) reactions were stopped at the same cycle during the log-linear phase of growth, determined by monitoring the reactions in real time (Fig. 2c, d). Products were taken at cycle X and cycle (X+2), where X is the cycle at which the RT-PCR product derived from experimental animals enters the log-linear growth phase. PCR products were visualized after 2% agarose gel electrophoresis. For statistical comparisons of Q-PCR data, cycles at which RT-PCR products from Ru486-induced and uninduced control samples entered the exponential phase were compared. The difference in PCR cycles required for identical levels of RT-PCR product (at early log-linear growth phase) from experimental and control animals is plotted in Fig. 2e, f. A one-cycle difference represents a twofold difference in starting template concentration. Each sample set was compared using the Student's *t*-test, and only results with a *P*-value <0.05 were considered statistically significant.

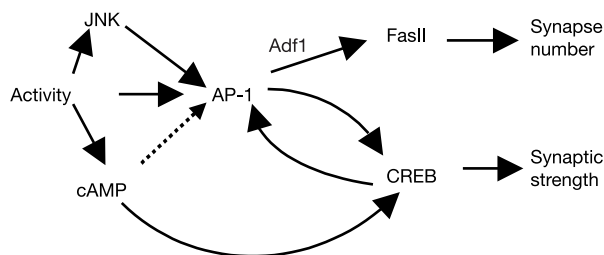


Figure 4 A model for plasticity regulation at the *Drosophila* larval neuromuscular synapse. AP-1 functions upstream of CREB to regulate plasticity. Adf1 influences bouton number but not synapse strength²⁷. AP-1 regulation by CREB is modelled as a positive feedback loop. An important cAMP-sensitive target (potentially CREB itself) functions downstream of AP-1 in regulation of synaptic strength. Upstream signals derive from neural activity-induced Ca²⁺ and cAMP signalling. The mechanism of a postulated cAMP-mediated activation of AP-1 is unknown; however, this study indicates the involvement of JNK signalling upstream of AP-1-dependent transcription.

Electrophysiology and data analysis

Synaptic currents were recorded in HL-3 saline from muscle 6 in segment A2 using a two-electrode voltage clamp²⁸. For EJC recordings, stimuli consisted of 0.2 ms pulses delivered at 1 Hz from an isolated pulse stimulator (AM systems 2100), gated by pClamp software (Axon Instruments). Intracellular glass microelectrodes were filled with 3 M KCl (voltage monitor electrode, resistance 7–10 MΩ) or saturated potassium citrate and 3 M KCl (current injection electrode, resistance 18–25 MΩ). Data were acquired from muscles clamped at –70 mV, using an Axoclamp 2B amplifier (Axon Instruments), and recorded with pClamp software. Data were filtered at 1 kHz before analysis. EJC amplitude was determined from 22–25 contiguous stimuli, and mEJC parameters were measured from one to two 41-s records, using the minianalysis software (Jaevin Software). Differences between means were compared using either *t*-tests (paired comparisons) or analysis of variance (ANOVA; multiple comparisons) in Microsoft Excel or SigmaPlot.

Received 26 November 2001; accepted 21 February 2002.

1. Bailey, C. H., Bartsch, D. & Kandel, E. R. Toward a molecular definition of long-term memory storage. *Proc. Natl Acad. Sci. USA* **93**, 13445–13452 (1996).
2. Curran, T. & Morgan, J. I. Fos: an immediate-early transcription factor in neurons. *J. Neurobiol.* **26**, 403–412 (1995).
3. Davis, G. W., Schuster, C. M. & Goodman, C. S. Genetic dissection of structural and functional components of synaptic plasticity. III. CREB is necessary for presynaptic functional plasticity. *Neuron* **17**, 669–679 (1996).
4. Sheng, M., McFadden, G. & Greenberg, M. E. Membrane depolarization and calcium induce *c-fos* transcription via phosphorylation of transcription factor CREB. *Neuron* **4**, 571–582 (1990).
5. Yin, J. C., Del Vecchio, M., Zhou, H. & Tully, T. CREB as a memory modulator: induced expression of a *Δ*CREB2 activator isoform enhances long-term memory in *Drosophila*. *Cell* **81**, 107–115 (1995).
6. Bartsch, D. *et al.* Aplysia CREB2 represses long-term facilitation: relief of repression converts transient facilitation into long-term functional and structural change. *Cell* **83**, 979–992 (1995).
7. Josselyn, S. A. *et al.* Long-term memory is facilitated by cAMP response element-binding protein overexpression in the amygdala. *J. Neurosci.* **21**, 2404–2412 (2001).
8. Sheng, M. & Greenberg, M. E. The regulation and function of *c-fos* and other immediate-early genes in the nervous system. *Neuron* **4**, 477–485 (1990).
9. Kelz, M. B. *et al.* Expression of the transcription factor deltaFosB in the brain controls sensitivity to cocaine. *Nature* **401**, 272–276 (1999).
10. Kockel, L., Homsy, J. G. & Bohmann, D. *Drosophila* AP-1: lessons from an invertebrate. *Oncogene* **20**, 2347–2364 (2001).
11. Davis, G. W. & Goodman, C. S. Genetic analysis of synaptic development and plasticity: homeostatic regulation of synaptic efficacy. *Curr. Opin. Neurobiol.* **8**, 149–156 (1998).
12. Zhong, Y., Budnik, V. & Wu, C. F. Synaptic plasticity in *Drosophila* memory and hyperexcitable mutants: role of cAMP cascade. *J. Neurosci.* **12**, 644–651 (1992).
13. Schuster, C. M., Davis, G. W., Fetter, R. D. & Goodman, C. S. Genetic dissection of structural and functional components of synaptic plasticity. Fasciclin II controls presynaptic structural plasticity. *Neuron* **17**, 641–654 (1996).
14. Bailey, C. H., Chen, M., Keller, F. & Kandel, E. R. Serotonin-mediated endocytosis of apCAM: an early step of learning-related synaptic growth in Aplysia. *Science* **256**, 645–649 (1992).
15. Eresh, S., Riese, J., Jackson, D. B., Bohmann, D. & Bienz, M. A CREB-binding site as a target for decapentaplegic signalling during *Drosophila* endoderm induction. *EMBO J.* **16**, 2014–2022 (1997).
16. Zeitlinger, J. *et al.* Defective dorsal closure and loss of epidermal decapentaplegic expression in *Drosophila* *fos* mutants. *EMBO J.* **16**, 7393–7401 (1997).
17. Karin, M., Liu, Z. & Zandi, E. AP-1 function and regulation. *Curr. Opin. Cell Biol.* **9**, 240–246 (1997).
18. Cheung, U. S., Shayan, A. J., Boulianne, G. L. & Atwood, H. L. *Drosophila* larval neuromuscular junction's responses to reduction of cAMP in the nervous system. *J. Neurobiol.* **40**, 1–13 (1999).
19. Yin, J. C. *et al.* Induction of a dominant-negative CREB transgene specifically blocks long-term memory in *Drosophila*. *Cell* **79**, 49–58 (1994).
20. Osterwalder, T., Yoon, K., White, B. & Keshishian, H. A conditional tissue-specific transgene system using inducible Gal4. *Proc. Natl Acad. Sci. USA* **98**, 12596–12601 (2001).
21. Chang, L. & Karin, M. Mammalian MAP kinase signalling cascades. *Nature* **410**, 37–40 (2001).
22. Riesgo-Escovar, J. R., Jenni, M., Fritz, A. & Hafen, E. The *Drosophila* Jun-N-terminal kinase is required for cell morphogenesis but not for DJun-dependent cell fate specification in the eye. *Genes Dev.* **10**, 2759–2768 (1996).
23. Yang, D. D. *et al.* Absence of excitotoxicity-induced apoptosis in the hippocampus of mice lacking the JNK3 gene. *Nature* **389**, 865–870 (1997).
24. Schwarzschild, M. A., Cole, R. L. & Hyman, S. E. Glutamate, but not dopamine, stimulates stress-activated protein kinase and AP-1-mediated transcription in striatal neurons. *J. Neurosci.* **17**, 3455–3466 (1997).
25. Sagasti, A. *et al.* The CaMKII α 43 activates the MAPKKK α 1 to execute a lateral signalling decision required for asymmetric olfactory neuron fates. *Cell* **105**, 221–232 (2001).
26. Martin-Blanco, E. *et al.* *puckered* encodes a phosphatase that mediates a feedback loop regulating JNK activity during dorsal closure in *Drosophila*. *Genes Dev.* **12**, 557–570 (1998).
27. DeZazzo, J. *et al.* *nalyot*, a mutation of the *Drosophila* myb-related Adf transcription factor, disrupts synapse formation and olfactory memory. *Neuron* **27**, 145–158 (2000).
28. Stewart, B. A., Atwood, H. L., Renger, J. J., Wang, J. & Wu, C. F. Improved stability of *Drosophila* larval neuromuscular preparations in haemolymph-like physiological solutions. *J. Comp. Physiol. A* **175**, 179–191 (1994).

Acknowledgements

We thank D. Bohmann and E. Hafen for strains, and T. Littleton and H. Bellen for anti-Syt antibodies. We also thank H. Keshishian and T. Osterwalder for sharing *elav*GS-GLA4 strains before publication; P. Etter for early assistance; D. Stimson for advice on electrophysiology; and P. Jansma for help with confocal microscopy. We thank P. Etter, R. Narayanan and members of the Ramaswami laboratory for useful discussions and/or comments on the manuscript. The work was funded by grants from the National Institute on Drug Abuse (primarily), the National Institute for Neurological Disorders and Stroke,

the Human Frontiers Science Program Organization, McKnight and Alfred P. Sloan Foundations to M.R., and NSF and NIH predoctoral fellowships to C.A.H.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.R. (e-mail: mani@u.arizona.edu).

Deafness and renal tubular acidosis in mice lacking the K-Cl co-transporter Kcc4

Thomas Boettger[†], Christian A. Hübner[†], Hannes Maier[‡], Marco B. Rust^{*}, Franz X. Beck[§] & Thomas J. Jentsch^{*}

^{*} Zentrum für Molekulare Neurobiologie, ZMNH, Universität Hamburg, Falkenried 94, 20246 Hamburg, Germany

[‡] HNO Klinik, Universitätsklinikum Eppendorf, Universität Hamburg, Martinistrasse 52, 20246 Hamburg, Germany

[§] Physiologisches Institut, Ludwig-Maximilians-Universität München, Pettenkoferstrasse 12, 80336 München, Germany

[†] These authors contributed equally to this work

Hearing depends on a high K⁺ concentration bathing the apical membranes of sensory hair cells. K⁺ that has entered hair cells through apical mechanosensitive channels is transported to the stria vascularis for re-secretion into the scala media¹. K⁺ probably exits outer hair cells by KCNQ4 K⁺ channels^{2,3}, and is then transported—by means of a gap junction system connecting supporting Deiters' cells and fibrocytes⁴—back to the stria vascularis. We show here that mice lacking the K⁺/Cl[−] (K-Cl) co-transporter Kcc4 (coded for by *Slc12a7*) are deaf because their hair cells degenerate rapidly after the beginning of hearing. In the mature organ of Corti, Kcc4 is restricted to supporting cells of outer and inner hair cells. Our data suggest that Kcc4 is important for K⁺ recycling^{4,5} by siphoning K⁺ ions after their exit from outer hair cells into supporting Deiters' cells, where K⁺ enters the gap junction pathway. Similar to some human genetic syndromes⁶, deafness in Kcc4-deficient mice is associated with renal tubular acidosis. It probably results from an impairment of Cl[−] recycling across the basolateral membrane of acid-secreting α -intercalated cells of the distal nephron.

Electroneutral K-Cl co-transporters have roles in cell volume regulation, transepithelial transport, and in the regulation of intracellular chloride concentration ([Cl]_i). Of the four mammalian K-Cl co-transporters (Kcc1–4), Kcc1 and Kcc3 are broadly expressed, whereas Kcc2 is neurone-specific⁷. In mice lacking Kcc2, the ensuing rise of neuronal [Cl]_i leads to excitatory responses to the normally inhibitory neurotransmitters GABA (γ -aminobutyric acid) and glycine, resulting in spasticity and early postnatal death⁸. Although the coupling of K⁺ to Cl[−] favours efflux and therefore lowers [Cl]_i under many circumstances, K-Cl co-transporters often operate near electrochemical equilibrium and may also mediate net ion uptake⁹.

We have now generated mice constitutively lacking Kcc4 (Supplementary Information Fig. 1a), an isoform with hitherto unknown physiological function. It is predominantly expressed in kidney, heart, lung and liver¹⁰. Western blots indicated that Kcc4 protein was absent in knockout mice (Supplementary Information Fig. 1c). *Kcc4*^{−/−} mice were born at the expected mendelian ratio

and were viable and fertile; however, their body weight was roughly 90% of that of their littermates.

Kcc4^{-/-} mice displayed no obvious abnormalities in activity and vision. Vestibular and motor function appeared normal on tilted plates and in rotarod experiments. The absence of acoustic Preyer reflexes in adult mice indicated severe hearing loss. Auditory brainstem responses were recorded from postnatal day 14 (P14) onwards, shortly after mice begin to hear. At this early age, hearing of *Kcc4*^{-/-} mice was normal. It quickly deteriorated during the following week after which mice were nearly deaf, with a hearing loss of 70–80 dB (Fig. 1). Histological analysis revealed that the inner ear developed normally and could not be distinguished from wild-type animals at P14 (Fig. 2a, b). At P21, however, outer hair cells (OHCs) of basal turns of the cochlea (which mediate high frequency hearing) were almost totally lost (Fig. 2c, d), whereas inner hair cells were still present. The degeneration proceeded from basal to apical turns (Supplementary Information Fig. 2). In adult knockout mice, the organ of Corti was lost completely in basal turns (Fig. 2f, h). In apical turns, some hair cells survived, accounting for the residual hearing ability in adult mice (Fig. 1). Probably as a consequence of the loss of hair cells¹¹, neurons of the cochlear ganglion degenerated as well (Fig. 2h). Even in adult knockout mice, however, there was no collapse of Reissner's membrane, which separates the scala media from the scala vestibuli (Fig. 2g, h). This contrasts with mice where inactivation of transport proteins of the stria vascularis impairs K⁺ secretion^{11–14}, suggesting that *Kcc4* is not essential for endolymph production.

Antibodies raised against *Kcc4* were used to determine its expression pattern (Fig. 3). At P8, before the onset of hearing, the antibodies stained membranes in the stria vascularis and in almost all cells of the organ of Corti (Fig. 3a). This included the developing epithelial and supporting cells, as well as the membranes of OHCs (Fig. 3c). Outer hair cells were identified by (green) staining for the K⁺ channel KCNQ4 (refs 2, 3), which was present in the entire basolateral membrane (Fig. 3c, d). At P14, when the organ has matured to allow hearing, KCNQ4 has reached its adult expression pattern in the basal membrane of OHCs³ (Fig. 3e, f). At this age, *Kcc4* was no longer expressed in hair cells and the stria, but was confined to a subset of supporting cells (Fig. 3b, e, g). Deiters' cells, which support OHCs at their basal pole and whose extensions are interposed between these, were labelled along their entire lateral membranes (Fig. 3e, g). *Kcc4* was also present in supporting cells of

inner hair cells (Fig. 3g).

K⁺ ions enter sensory hair cells from the K⁺-rich endolymph through apical mechanosensitive channels. They exit from OHCs presumably through KCNQ4 K⁺ channels^{2,3} (Fig. 3j). After leaving OHCs, K⁺ must be removed, partially by uptake into Deiters' cells. In the potassium recycling model of the inner ear^{1,5,15–18}, K⁺ then diffuses through a gap junction system connecting Deiters' cells to adjacent epithelial cells. After passing through a distinct fibrocyte gap junction system, it then contributes to the K⁺ supply for the secretory stria vascularis (Fig. 3j).

The mechanism of K⁺ uptake into Deiters' cells has remained unclear. These cells express only very low levels of Na/K-ATPase^{18,19}. Their membrane conductance is dominated by a K⁺ channel²⁰. Owing to its outward rectification, however, the K⁺ channel is closed at resting voltages²⁰, and hence cannot mediate K⁺ uptake. Our study suggests that K⁺ uptake into Deiters' cells instead occurs through *Kcc4*. This resembles K⁺ uptake by neuronal K-Cl co-transporters, which can buffer the extracellular K⁺ concentration ([K⁺]_o) that can rise during neuronal activity⁹. Similarly, [K⁺]_o rises in the tunnel of Corti during sound exposure²¹, an effect expected to be even more pronounced in the narrow space between OHCs and Deiters' cells. Unlike active pumping by the Na/K-ATPase, K⁺ uptake through *Kcc4* would be energetically 'cheap' and would occur close to electrochemical equilibrium. Neither the inflow of K⁺ into OHCs through their apical transduction channels, nor its basal efflux through KCNQ4 or its removal by *Kcc4* would directly

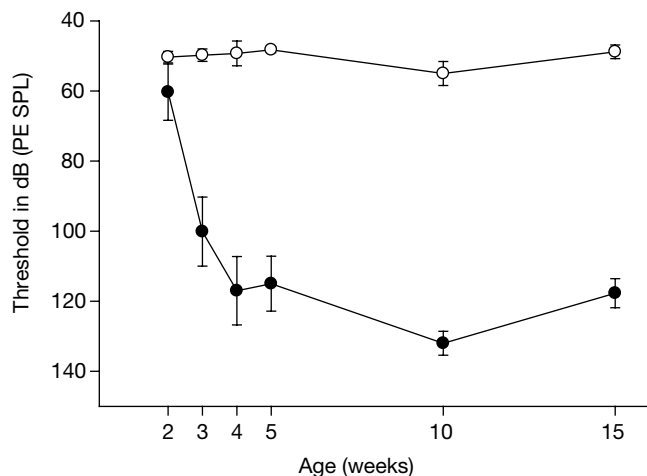


Figure 1 Hearing loss of *Kcc4*^{-/-} mice. Thresholds in peak equivalent sound pressure level (PE SPL; re 20 μ Pa) of auditory brain stem responses to clicks in wild-type or heterozygous (open circles) and knockout (filled circles) mice of different ages. Vertical bars indicate the standard error of five or more animals.

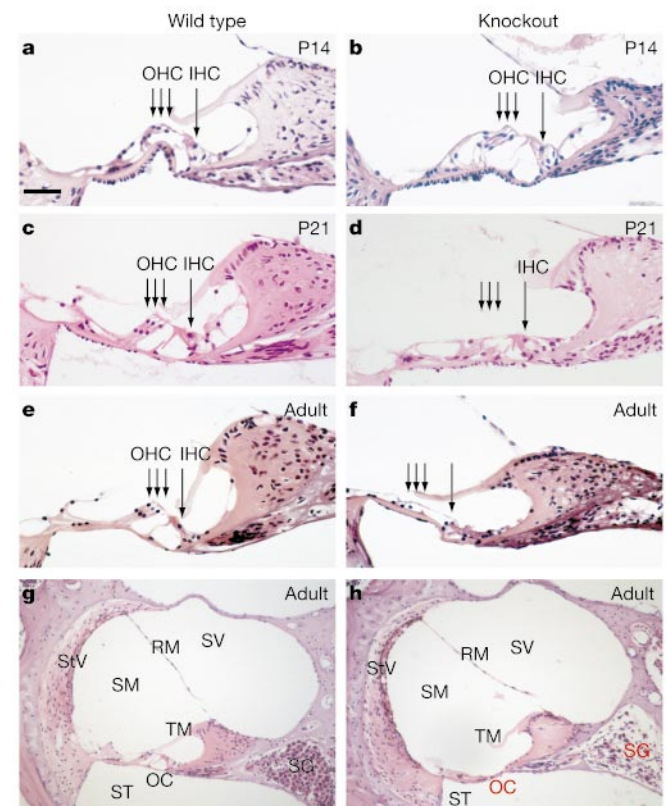


Figure 2 Inner ear morphology by haematoxylin/eosin staining. **a, b**, At postnatal day 14 (P14), there is no difference in the organ of Corti between wild-type (**a**) and knockout (**b**) mice. **c, d**, At P21, outer hair cells (OHCs) are lost in knockout mice (**d**). **e–h**, Adult (3 months of age) cochlea turns reveal normal stria vascularis (SV) and Reissner's membrane (RM), but a loss of OHCs and neuronal degeneration in the spiral ganglion (SG) in the knockout (**f, h**) compared with the wild type (**e, g**). Three animals per age and genotype gave similar results. IHC, inner hair cell; OC, organ of Corti; SM, scala media; ST, scala tympani; SV, scala vestibuli; TM, tectorial membrane.

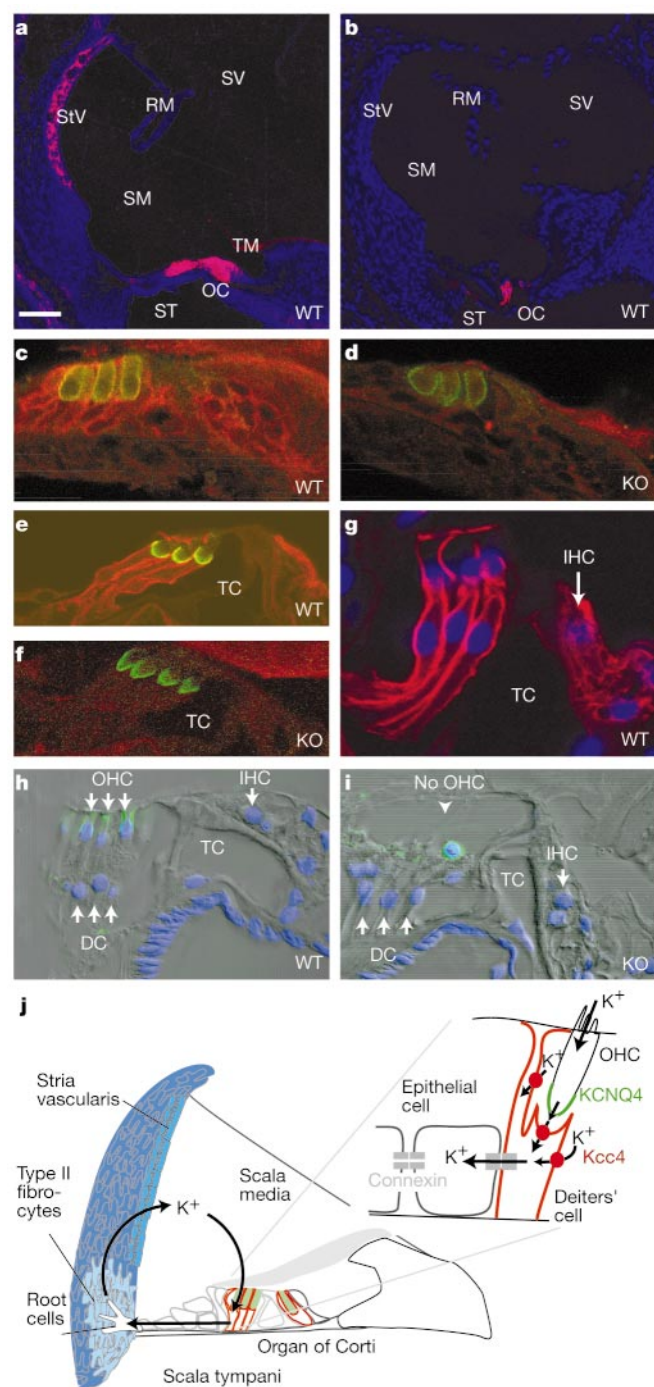


Figure 3 *Kcc4* in the cochlea. Abbreviations are the same as in Fig. 2. **a, b**, Confocal images of cochlear cross-sections at postnatal day 8 (P8) (**a**) and adult age (**b**) stained for *Kcc4* (in red) in the wild type (WT). Nuclei are stained in blue (TOTO-3). **c**, Higher magnification of the organ of Corti at P8 reveals widespread *Kcc4* expression. **d**, *Kcc4* is absent in the knockout mouse (KO). *KCNQ4* (green stain) labels basolateral OHC membranes, which co-stain (yellow) for *Kcc4* in the wild type (**c**). **e, f**, Cochlea of wild-type (**e**) and knockout (**f**) mice at P14 stained for *Kcc4* (red) and *KCNQ4* (green). **g**, In the adult wild-type cochlea, *Kcc4* (red) is restricted to Deiters' cells and the phalangeal cell enveloping the inner hair cell (nuclei; in blue). **h, i**, Organ of Corti in P17 wild-type (**h**) and knockout (**i**) mice stained with the OHC marker prestin (green)²³. Nuclei are stained blue. Outer hair cells are intact in the organ of Corti in wild-type mice, but are mostly lost in knockout mice. An OHC nucleus left in the organ of Corti of a knockout mouse is surrounded by a shrunken membrane containing prestin (arrowhead). Note intact Deiters' cells. The scale bar in **a** corresponds to 53 μm in **a, b**; 12 μm in **c**; 11 μm in **d, e**; 25 μm in **f**; 7 μm in **g**; and 13 μm in **h, i**. **j**, Model for inner ear K^+ recycling^{1,5}. TC, tunnel of Corti.

require metabolic energy. This makes biological sense, as placing the main energy source into the stria vascularis takes metabolic stress away from sensory cells.

The loss of *Kcc4* is expected to impede the removal of K^+ ions that have left OHCs. This would change the ionic composition of the small extracellular space that surrounds the basolateral membranes of OHCs, eventually leading to their degeneration. Similarly, mutations in connexin proteins²² are thought to cause deafness by blocking K^+ removal further downstream at the gap junction level¹. Deafness caused by *KCNQ4* mutations was attributed to chronic K^+ overload^{2,3} by impaired K^+ efflux from OHCs. Importantly, OHCs of *Kcc4*^{-/-} mice degenerated before Deiters' cells were lost (Fig. 3h, i), although Deiters' cells and not OHCs normally express *Kcc4* at this stage (Fig. 3e, g). This is consistent with a disturbance of extracellular homeostasis due to impaired salt uptake by Deiters' cells, and may lead to death of OHCs by osmotic stress or membrane depolarization. Furthermore, the degeneration began after the onset

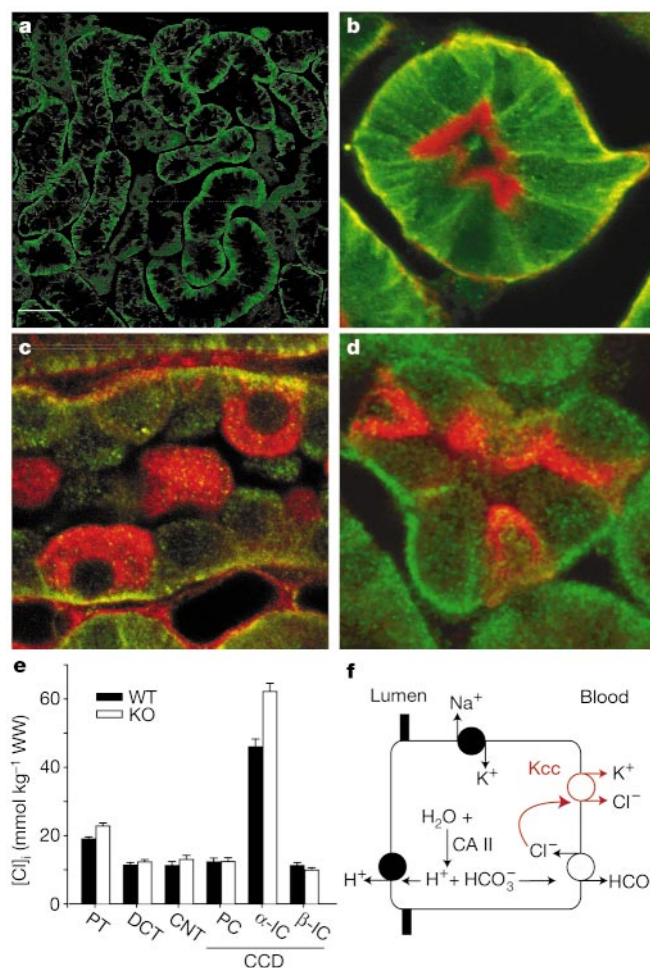


Figure 4 *Kcc4* in the kidney. **a**, Overview of *Kcc4*-stained (green) kidney cortex. **b**, *Kcc4* (green) in proximal tubules. Brush borders are stained red by phalloidin. **c**, α -Intercalated cells (identified by red H^+ -ATPase staining) of the cortical collecting duct (CCD) express *Kcc4* (green). **d**, Basolateral staining of *Kcc4* (green) in CCD intercalated cells (which lack aquaporin 2 (red)). The scale bar in **a** corresponds to 31 μm in **a**; 6 μm in **b**; and 5 μm in **c, d**. **e**, X-ray microprobe analysis of renal chloride concentrations. In *Kcc4*^{-/-} mice, $[\text{Cl}]_i$ (mmol kg^{-1} wet weight) is significantly elevated over wild type in CCD α -intercalated cells (α -ICs) and in proximal tubular (PT) cells. No significant difference was found in distal convoluted tubules (DCT), connecting tubules (CNT), nor in CCD principal cells (PC) or β -intercalated cells (β -ICs). Na^+ and K^+ concentrations were unchanged. Greater than 11 cells of each type were measured (from three animals). WT, wild type; KO, knockout. **f**, Proposed role of *Kcc4* in acid secretion by α -intercalated cells.

of hearing, possibly related to the increased K^+ influx through sensory OHCs. Cl^- gradients are important determinants of Kcc4 function, but unfortunately Cl^- transport in the organ of Corti is poorly understood. As crucial aspects of OHC function depend on chloride²³, one may speculate that changes in $[Cl^-]_o$ might contribute to compromising OHC function in *Kcc4*^{-/-} mice.

As in some forms of human syndromic hearing loss⁶, deafness in *Kcc4*^{-/-} mice was associated with renal tubular acidosis. The urine of knockout mice was more alkaline than in wild-type littermates (pH 7.3 ± 0.1 (knockout) compared with pH 6.4 ± 0.1 (wild type); $n = 7$, $P < 0.001$), whereas concentrations of Na^+ , K^+ and Cl^- were not changed. Consistent with a defect in urinary acidification, blood gas analysis indicated a compensated metabolic acidosis with significantly decreased base excess. Immunofluorescence revealed that Kcc4 is expressed in basolateral membranes of several nephron segments (Fig. 4a–d). Of note, this included α -intercalated cells (Fig. 4c, d). The apical proton ATPase of α -intercalated cells secretes H^+ into the lumen of the distal nephron (Fig. 4f). At the basolateral membrane, acid equivalents are transported by the anion exchanger AE1 (ref. 24). As basolateral HCO_3^- efflux is coupled to Cl^- uptake, Kcc4 may be required for basolateral Cl^- extrusion. Energy-dispersive X-ray microanalysis was used to compare $[Cl^-]$ in renal cells of wild-type and knockout mice (Fig. 4e). $[Cl^-]$ was unchanged in principal cells and in the distal convoluted tubule, which express Kcc4 at low or undetectable levels. By contrast, it was increased in proximal tubules and particularly in α -intercalated cells of knockout mice. High basal $[Cl^-]$ in α -intercalated cells of wild-type mice was described in previous microprobe studies²⁵, but it should be kept in mind that electron microprobe analysis determines total Cl, not Cl^- activity. Considering the prominent Cl^-/HCO_3^- exchange activity in α -intercalated cells (Fig. 4f), the rise in $[Cl^-]$ predicts a more alkaline intracellular pH in the knockout mouse. This will decrease apical H^+ secretion by increasing the electrochemical gradient against which pumping has to occur. Thus, Kcc4 joins the H^+ -ATPase⁶ and the AE1 anion exchanger²⁶ as the third transport protein of α -intercalated cells whose mutation entails renal tubular acidosis.

This work has revealed important and unexpected physiological roles of Kcc4 in the inner ear K^+ transport pathway and in renal acidification. By removing K^+ from the fluid around OHCs, Kcc4 is essential for maintaining the integrity of these sensory cells. In the kidney, Kcc4 is crucial for Cl^- extrusion across the basolateral membrane of acid-secreting α -intercalated cells. As a consequence, the loss of Kcc4 leads to deafness associated with renal tubular acidosis. □

Methods

Disruption of *Kcc4*

Mouse genomic *Kcc4* clones were from a 129/SV λ library (Stratagene). MPI2 mouse embryonic stem cells²⁷ were electroporated with a vector to introduce a neomycin cassette flanked by *loxP* sites (floxed), and a third *loxP* site. Recombinant clones were transfected with a plasmid expressing Cre-recombinase to remove the two exons coding for the first two transmembrane spans of Kcc4 (see Supplementary Information Fig. 1). Mouse lines were established from two independent embryonic stem cell clones. Studies were performed in a mixed 129SV (ref. 27)/C57Bl6 background. We used appropriate littermates as controls.

Antibodies

Polyclonal antibodies were raised against a peptide sequence (EKHRNKDTGPSGFKDC) from the Kcc4 carboxy terminus. Two rabbit sera and one guinea-pig serum proved useful in western blotting and immunofluorescence—they gave identical results. Their specificity was confirmed by analysis of *Kcc4*^{-/-} tissues (Fig. 3c–f; see also Supplementary Information Figs 1c and 4). We also used polyclonal antisera against KCNQ4 (ref. 3), prestin and aquaporin 2 (ref. 28), and the monoclonal antibody E11 against the H^+ -ATPase.

Morphology

Mouse kidney and inner ear sections were prepared and analysed by haematoxylin/eosin staining and immunofluorescence as described²⁹.

Electron microprobe analysis

Kidneys were shock frozen in an isopentane/propane mixture (1:4) at $-196^\circ C$. One-micrometre sections were cut on an ultracryomicrotome at $-90^\circ C$, and freeze dried. Element concentrations were determined using a scanning transmission electron microscope fitted to an X-ray detector system as described³⁰.

Hearing tests

Auditory evoked brain-stem responses (ABR) to clicks were recorded in mice anaesthetized with Rompun/Ketanest. Acoustic stimulation and recording of evoked potentials used an Evokelect system (Pilot Blankenfelde). Bioelectric potentials were recorded by subcutaneous silver electrodes at the vertex (reference), forehead (ground), and ventrolateral to the stimulated ear. Acoustic click stimuli were delivered mono-aurally using a Beyer DT-48 earphone, and were monitored with a probe microphone (MK301, Microtech Gefell) integrated into the earpiece. Calibration was done in a 19 μ l volume using a second probe microphone (Brüel & Kjær 4135) with a sound-level meter (Brüel & Kjær 2215). Click stimuli had a main component of approximately 200 μ s duration and a flat spectra (± 5 dB) with an upper corner frequency of 5.5 kHz. Alternating clicks were applied at a rate of 21 per s and averaged 400–2,000 times.

Urine collection

Mice were kept singly in mouse metabolic cages (PHYMED) with free access to water and chow. After adaptation to the cage for three days, urine was collected over 24 h periods.

Received 14 December 2001; accepted 15 February 2002.

- Steel, K. P. & Kros, C. J. A genetic approach to understanding auditory function. *Nature Genet.* **27**, 143–149 (2001).
- Kubisch, C. et al. KCNQ4, a novel potassium channel expressed in sensory outer hair cells, is mutated in dominant deafness. *Cell* **96**, 437–446 (1999).
- Kharkovets, T. et al. KCNQ4, a K^+ channel mutated in a form of dominant deafness, is expressed in the inner ear and the central auditory pathway. *Proc. Natl Acad. Sci. USA* **97**, 4333–4338 (2000).
- Kikuchi, T., Kimura, R. S., Paul, D. L., Takasaka, T. & Adams, J. C. Gap junction systems in the mammalian cochlea. *Brain Res. Brain Res. Rev.* **32**, 163–166 (2000).
- Jentsch, T. J. Neuronal KCNQ potassium channels: physiology and role in disease. *Nature Rev. Neurosci.* **1**, 21–30 (2000).
- Karet, F. E. et al. Mutations in the gene encoding B1 subunit of H^+ -ATPase cause renal tubular acidosis with sensorineural deafness. *Nature Genet.* **21**, 84–90 (1999).
- Lauf, P. K. & Adragna, N. C. K-Cl cotransport: properties and molecular mechanism. *Cell Physiol. Biochem.* **10**, 341–354 (2000).
- Hübner, C. A. et al. Disruption of KCC2 reveals an essential role of K-Cl cotransport already in early synaptic inhibition. *Neuron* **30**, 515–524 (2001).
- Payne, J. A. Functional characterization of the neuronal-specific K-Cl cotransporter: implications for $[K^+]_o$ regulation. *Am. J. Physiol.* **273**, C1516–C1525 (1997).
- Mount, D. B. et al. Cloning and characterization of KCC3 and KCC4, new members of the cation-chloride cotransporter gene family. *J. Biol. Chem.* **274**, 16355–16362 (1999).
- Vetter, D. E. et al. Inner ear defects induced by null mutation of the *isk* gene. *Neuron* **17**, 1251–1264 (1996).
- Delpire, E., Lu, J., England, R., Dull, C. & Thorne, T. Deafness and imbalance associated with inactivation of the secretory Na-K-2Cl co-transporter. *Nature Genet.* **22**, 192–195 (1999).
- Dixon, M. J. et al. Mutation of the Na-K-Cl co-transporter gene *Slc12a2* results in deafness in mice. *Hum. Mol. Genet.* **8**, 1579–1584 (1999).
- Lee, M. P. et al. Targeted disruption of the *Kvlqt1* gene causes deafness and gastric hyperplasia in mice. *J. Clin. Invest.* **106**, 1447–1455 (2000).
- Kikuchi, T., Kimura, R. S., Paul, D. L. & Adams, J. C. Gap junctions in the rat cochlea: immunohistochemical and ultrastructural analysis. *Anat. Embryol. (Berl.)* **191**, 101–118 (1995).
- Spicer, S. S. & Schulte, B. A. Evidence for a medial K^+ recycling pathway from inner hair cells. *Hear Res.* **118**, 1–12 (1998).
- Wada, J., Kambayashi, J., Marcus, D. C. & Thalmann, R. Vascular perfusion of the cochlea: effect of potassium-free and rubidium-substituted media. *Arch. Otorhinolaryngol.* **225**, 79–81 (1979).
- Weber, P. C., Cunningham, C. D. III & Schulte, B. A. Potassium recycling pathways in the human cochlea. *Laryngoscope* **111**, 1156–1165 (2001).
- Spicer, S. S. & Schulte, B. A. Creatine kinase in epithelium of the inner ear. *J. Histochem. Cytochem.* **40**, 185–192 (1992).
- Nenov, A. P., Chen, C. & Bobbin, R. P. Outward rectifying potassium currents are the dominant voltage activated currents present in Deiter's cells. *Hear Res.* **123**, 168–182 (1998).
- Johnstone, B. M., Patuzzi, R., Syka, J. & Sykova, E. Stimulus-related potassium changes in the organ of Corti of guinea-pig. *J. Physiol.* **408**, 77–92 (1989).
- Kelsell, D. P. et al. Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. *Nature* **387**, 80–83 (1997).
- Oliver, D. et al. Intracellular anions as the voltage sensor of prestin, the outer hair cell motor protein. *Science* **292**, 2340–2343 (2001).
- Rodríguez-Soriano, J. New insights into the pathogenesis of renal tubular acidosis—from functional to molecular studies. *Pediatr. Nephrol.* **14**, 1121–1136 (2000).
- Beck, F. X., Dörge, A., Giebisch, G. & Thürk, K. Renal excretion of rubidium and potassium: an electron microprobe and clearance study. *Kidney Int.* **34**, 455–462 (1988).
- Bruce, L. J. et al. Familial distal renal tubular acidosis is associated with mutations in the red cell anion exchanger (Band 3, AE1) gene. *J. Clin. Invest.* **100**, 1693–1707 (1997).
- Voss, A. K., Thomas, T. & Gruss, P. Germ line chimeras from female ES cells. *Exp. Cell Res.* **230**, 45–49 (1997).
- Nielsen, S., DiGiovanni, S. R., Christensen, E. I., Knepper, M. A. & Harris, H. W. Cellular and subcellular immunolocalization of vasopressin-regulated water channel in rat kidney. *Proc. Natl Acad. Sci. USA* **90**, 11663–11667 (1993).
- Estévez, R. et al. Barttin is a Cl^- channel β -subunit crucial for renal Cl^- reabsorption and inner ear K^+ secretion. *Nature* **414**, 558–561 (2000).

30. Beck, F. X., Schmolke, M., Guder, W. G., Dörge, A. & Thurauf, K. Osmolytes in renal medulla during rapid changes in papillary tonicity. *Am. J. Physiol.* **262**, F849–F856 (1992).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank J. Faulhaber and H. Ehmke for blood gas analysis; M. Bösl for blastocyst injection and implantation; M. Knipper for the prestin antiserum; M. Knepper for the aquaporin 2 antiserum; S. Gluck for the proton ATPase antiserum; M. Kolster, B. Dierkes and I. Öztürk for technical assistance; H. Voss for taking care of animals; and U. Koch for support. This work was supported by grants from the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie, and the Prix Louis-Jeanet de Médecine to T.J.J.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.J.J. (e-mail: jentsch@zmn.uni-hamburg.de).

Structural determinants for GoLoco-induced inhibition of nucleotide release by G α subunits

Randall J. Kimple*, Michelle E. Kimple†, Laurie Betts*†, John Sondek*‡§ & David P. Siderovski*‡§

* Department of Pharmacology; † Department of Biochemistry and Biophysics; ‡ Lineberger Comprehensive Cancer Center; and § UNC Neuroscience Center, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA

Heterotrimeric G-proteins bind to cell-surface receptors and are integral in transmission of signals from outside the cell. Upon activation of the G α subunit by binding of GTP, the G α and G $\beta\gamma$ subunits dissociate and interact with effector proteins for signal transduction. Regulatory proteins with the 19-amino-acid GoLoco motif^{1,2} can bind to G α subunits and maintain G-protein subunit dissociation in the absence of G α activation^{3–7}. Here we describe the structural determinants of GoLoco activity as revealed by the crystal structure of G α_{i1} -GDP bound to the GoLoco region of the 'regulator of G-protein signalling' protein

RGS14. Key contacts are described between the GoLoco motif and G α protein, including the extension of GoLoco's highly conserved Asp/Glu-Gln-Arg triad into the nucleotide-binding pocket of G α to make direct contact with the GDP α - and β -phosphates. The structural organization of the GoLoco-G α_{i1} complex, when combined with supporting data from domain-swapping experiments, suggests that the G α all-helical domain and GoLoco-region carboxy-terminal residues control the specificity of GoLoco-G α interactions.

In heterotrimeric G-protein signalling, cell surface receptors (GPCRs) are coupled to membrane-associated heterotrimers comprising a GTP-hydrolysing G α subunit and a G β -G γ dimer. G $\beta\gamma$ binds tightly to GDP-bound G α , enhancing the coupling of G α to the receptor and acting as a guanine nucleotide dissociation inhibitor (GDI) to inhibit spontaneous GDP release^{8,9}. Agonist-promoted exchange of bound GDP for GTP alters the conformation of three G α 'switch' regions (I–III) and allows G $\beta\gamma$ dissociation and subsequent effector interactions by both α -GTP and free G $\beta\gamma$. Intrinsic, or RGS protein-accelerated¹⁰, GTP hydrolysis by G α returns the subunit to the GDP-bound state and allows G $\alpha\beta\gamma$ re-assembly and termination of effector interactions. GoLoco-motif proteins interact specifically with GDP-bound G $\alpha_{i/o}$ -class G α subunits, preventing both GDP release^{3–6} and G $\beta\gamma$ re-assembly^{6,7}, and thus permitting continued G $\beta\gamma$ -effector interactions in the absence of G α activation². The GoLoco motif is present in RGS12, RGS14, LOCO, Purkinje-cell protein-2 (Pcp2) and Rap1GAP isoforms, and is repeated in tandem arrays within the *Drosophila* protein Rapsynoid (also known as Partner of Inscuteable, Pins) and its mammalian homologues AGS3 and LGN (ref. 1). During development of the *Drosophila* nervous system, the binding of Pins to G α_s and the resultant displacement of G $\beta\gamma$, is thought to underlie mitotic spindle re-orientation, which is critical for the asymmetric cell divisions exhibited by embryonic neuroblasts and sensory organ precursor cells⁷. More recently, the Pins-related protein LGN has been shown to be essential for mitotic spindle assembly and organization in mammalian cells¹¹.

To ascertain the structural determinants of selective α -GDP binding and novel GDI activity exhibited by GoLoco proteins, we determined the crystal structure of the RGS14 GoLoco region bound to an adenylyl cyclase-inhibitory G α subunit (α_{i1} -GDP). Diffraction data collected from a single crystal at 100 K were used to refine the structure to 2.7 Å resolution (Supplementary Information Table A) using the structure of α_{i1} -GDP-Mg²⁺ as a model for molecular replacement¹².

Of the 36 amino acids of the R14GL peptide (residues r496–r530, numbered according to full-length rat RGS14), 35 are ordered

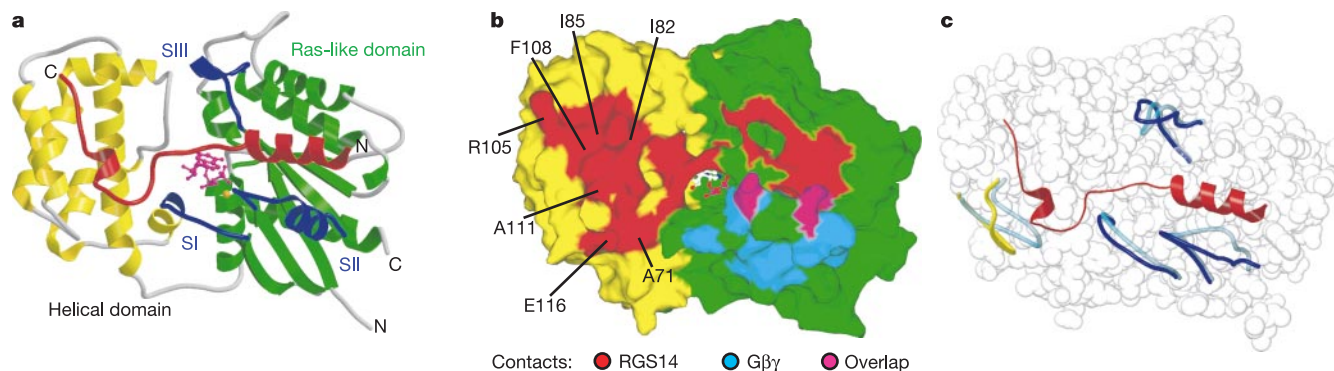


Figure 1 α_{i1} -GDP in complex with the RGS14 GoLoco region. **a**, Ribbon drawing of R14GL peptide (red) in contact with the Ras-like (green) and all-helical (yellow) domains of G α_{i1} . Also shown are the three switch regions of G α_{i1} (blue), GDP (magenta) and Mg²⁺ (orange). **b**, Molecular surface of R14GL (red) and G $\beta\gamma$ (cyan) contacts on G α_{i1} -GDP, denoting

shared switch II residue contacts (magenta). Highlighted are G α_{i1} residues that contact the R14GL peptide and are different within G α_o . **c**, Space fill model of α_{i1} -GDP in its R14GL-bound conformation (switch regions in blue, α B- α C loop in yellow) and G $\beta\gamma$ -bound conformation (in cyan).

within the complex (Fig. 1a and Supplementary Information); more than half of the total surface area of the peptide is buried by the α_{i1} subunit (1,900 Å² buried out of 3,666 Å² total surface area; Supplementary Information Table B). The amino terminus of the R14GL peptide (r496–r508) forms an α -helix that is sandwiched between switch II and the $\alpha 3$ helix of the α_{i1} Ras-like domain (Fig. 1a); contacts between R14GL and switch II residues overlap with contacts made between α_{i1} ·GDP and $G\beta\gamma$ subunits (Fig. 1b). The side chain of Gln r508, which caps the α -helix C terminus, is extensively buried within α_{i1} . Contacts made between α_{i1} and the N-terminal portion of the R14GL peptide (residues r496–r516) all involve residues that are identical between α_{i1} and α_o . The GoLoco peptide continues across the interdomain region between the Ras-like and all-helical domains in a more random conformation; in this region, conserved residues of the 19-amino-acid GoLoco motif contact both switch I and the bound guanine nucleotide. Surprisingly, the R14GL peptide C terminus (residues r517–r530), a region poorly conserved among GoLoco proteins, interacts extensively (660 Å² buried surface area) with the αA and αB helices of the α_{i1} α -helical domain (Fig. 1a). More surprisingly, $G\alpha$ residues within this region that contact the GoLoco peptide (Ala 71, Ile 82, Ile 85, Arg 105, Phe 108, Ala 111 and Glu 116; Fig. 1b) are the only contact residues that differ between α_{i1} and α_o subunits.

Interactions with the R14GL peptide alter the conformationally flexible switches I and II relative to uncomplexed α_{i1} ·GDP·Mg²⁺ (not shown) and $G\beta\gamma$ -bound α_{i1} ·GDP·Mg²⁺ (Fig. 1c). In particular, the significant deformation of switch II (for example, Arg 208 moves about 6 Å relative to heterotrimer configuration) would preclude coincident $G\beta\gamma$ binding to GoLoco-complexed α_{i1} ·GDP (Fig. 1c). This supports previous findings that $G\beta\gamma$ - and GoLoco-binding are mutually exclusive^{2,6,7}. The conformation of switch region III is also altered relative to the heterotrimer structure (Fig. 1c), and the αB – αC loop of the α_{i1} helical domain, previously dubbed ‘switch IV’¹³, is displaced away from the Ras-like domain (for example, Ala 114 moves about 11 Å, and Glu 116 moves about 4 Å relative to the heterotrimer configuration) to accommodate

binding of the GoLoco C-terminal region.

The highly conserved Asp–Gln–Arg triad within the GoLoco motif participates directly in GDP binding by extending the arginine side chain into the nucleotide binding pocket (Fig. 2), highly reminiscent of the catalytic ‘arginine finger’ employed by GTPase-accelerating proteins (GAPs) for Ras-superfamily GTPases^{14–16}. Gln r515 is positioned away from the GDP binding pocket and makes both side-chain and backbone interactions with Asp r514 of the peptide and Gln 147 and Asn 149 of the $G\alpha$ subunit; these constraints placed on the Gln r515 conformation position the peptide to allow full insertion of the neighbouring Arg r516 residue into the GDP binding pocket. The guanidinium group of Arg r516 contacts the α - and β -phosphates and the bridging oxygen between them (Fig. 2a). The role of Arg r516 in the GoLoco– $G\alpha$ interaction is underscored by the diminished GDI activity seen on mutation of Arg r516 to alanine (R516A) or leucine (R516L) (nearly a tenfold reduction in activity; Fig. 2b), or the complete loss of GDI activity when changed to phenylalanine¹⁷. Although the nonconservative replacement of arginine with the bulky, hydrophobic phenylalanine also results in greatly decreased binding affinity^{2,17}, neither the R516L nor the R516A mutation decreased the affinity of the GoLoco– α_{i1} ·GDP interaction. Dissociation constants (K_D) derived by a surface plasmon resonance biosensor for both mutants were comparable to that previously obtained for the wild-type RGS14 GoLoco region: for glutathione S-transferase (GST)–RGS14_{496–513} [R516A], $K_D = 52$ nM ($\chi^2 = 5.2$); and for GST–RGS14_{496–513} [R516L], $K_D = 70$ nM ($\chi^2 = 31.9$); compared with wild-type GST–RGS14_{496–513} for which $K_D = 65$ nM ($\chi^2 = 3.7$)⁵. Conservative mutation of a nucleotide-interacting arginine residue has been known to diminish biochemical activity without loss of binding affinity. Conservative mutations to the catalytic arginine finger of several GAPs, including p120-RasGAP¹⁴, NF1/neurofibromin¹⁵, RhoGAP¹⁶, and eIF5 (ref. 18), were previously reported to affect the ability to stimulate GTP hydrolysis without influencing complex formation with cognate GTPase partners.

Arg 178 within switch I of the α_{i1} subunit is critically involved in GTP hydrolysis by stabilizing the γ -phosphate leaving group^{19–21}. In

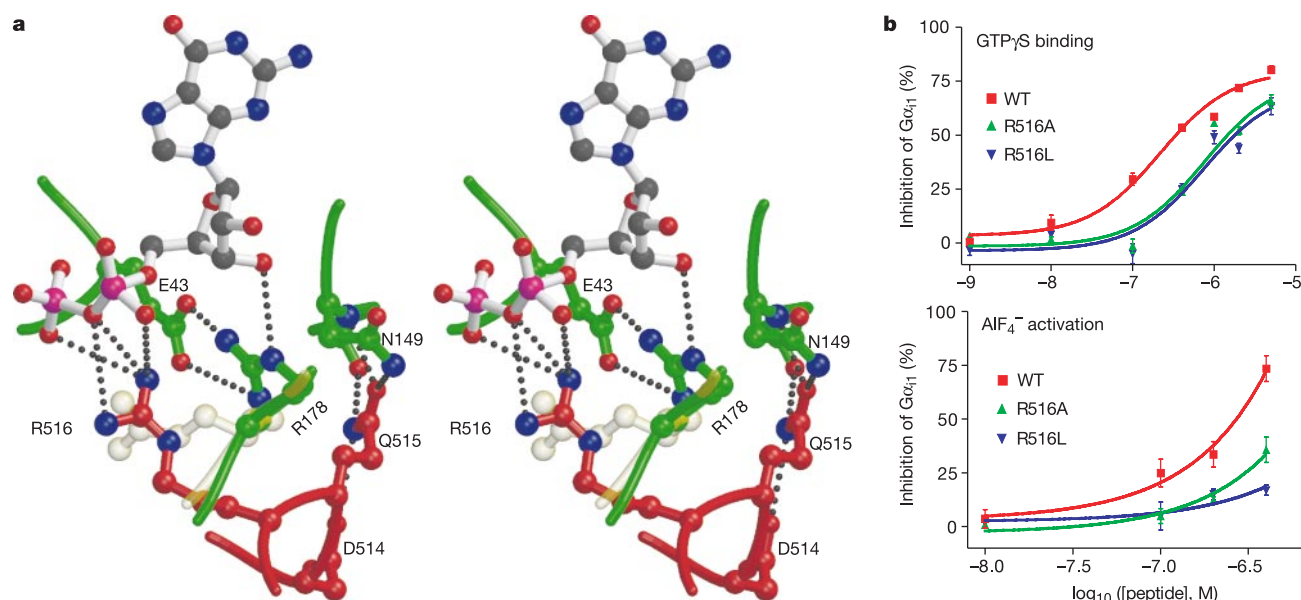


Figure 2 Role of the GoLoco motif Asp–Gln–Arg triad in GDI activity. **a**, Stereo view of the remodeled nucleotide binding pocket. The $G\alpha_{i1}$ Arg 178 side chain (transparent yellow) is re-oriented (green) to form a salt bridge with Glu 43 of the α_{i1} subunit, thus allowing Arg r516 of the GoLoco motif (red) to approach and interact with the α - and β -phosphate oxygens (and bridging oxygen) of GDP. **b**, GST–RGS14 GoLoco fusion proteins (GST–

RGS14_{496–531}) bearing alanine (R516A) or leucine (R516L) substitutions show decreased GDI activity relative to the wild-type GST–RGS14_{496–531} (WT), as measured both by BODIPY–GTPγS binding (left) and by AIF₄⁻-induced increase of intrinsic tryptophan fluorescence (right).

uncomplexed α_{i1} -GDP, the Arg 178 guanidinium group contributes to GDP binding by forming hydrogen bonds with α - and β -phosphate oxygens¹³. Within GoLoco-bound α_{i1} -GDP, Arg 178 is displaced from its phosphate contacts by a salt bridge with Glu 43 of the $G\alpha$ subunit P-loop, resulting in a hydrogen bond between the side-chain guanidinium group and the 3' hydroxyl of the GDP ribose (Fig. 2a). This pairing of Arg 178 and Glu 43 is also induced by the binding of $G\beta_1$ in the heterotrimer complex (α_{i1} -GDP- $G\beta_1$ - $G\gamma_2$; ref. 22) and is absent in uncomplexed α_{i1} -GDP. No additional GDP-binding residues are observed to undergo displacement greater than 1 Å in GoLoco-bound α_{i1} -GDP relative to the uncomplexed α_{i1} -GDP structure. We propose that the creation of new contacts to the bound guanosine diphosphate by remodelling of Arg 178 and addition of Arg 516 to the GDP binding pocket represent principal aspects of the molecular basis of GoLoco GDI activity.

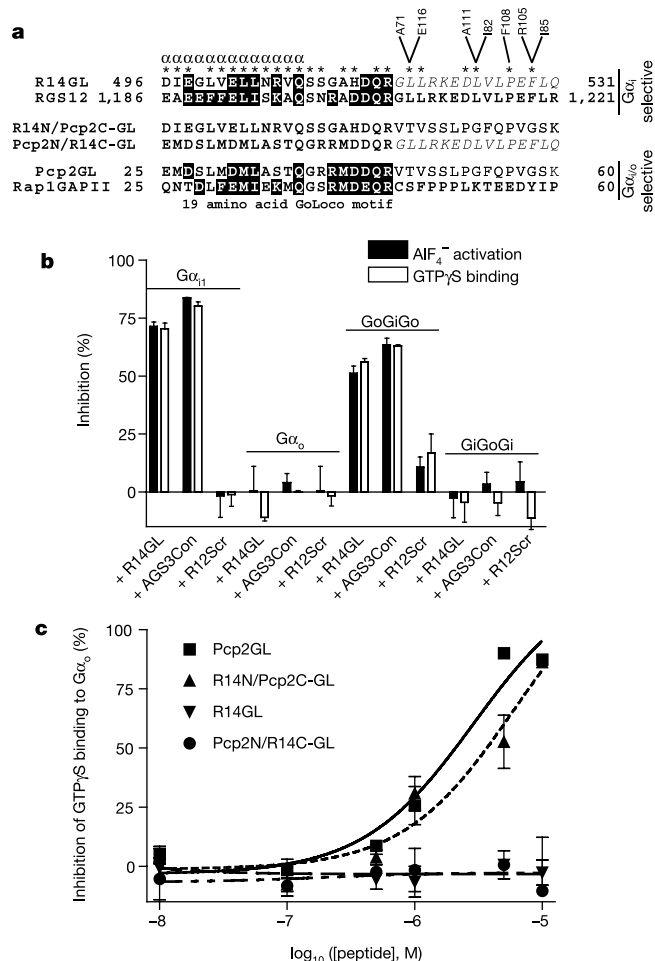


Figure 3 Roles of GoLoco C terminus and $G\alpha$ all-helical domain in GDI selectivity. **a**, Alignment of GoLoco regions for which $G\alpha$ specificity is known. The core GoLoco motif is boxed. N-terminal α -helical residues are denoted by ' α '. Asterisks denote RGS14- $G\alpha$ contacts. $G\alpha_{i1}$ -specific contacts are identified by connecting lines. Residues C-terminal to the 19-amino-acid GoLoco motif are swapped within chimaeric peptides as indicated. **b**, $G\alpha_{i1}$ -selective GDI activity is dependent on the $G\alpha$ all-helical domain, as assessed by GTP γ S binding (open bars) and AIF₄⁻ induced intrinsic fluorescence (filled bars). 'AGS3Con' denotes pre-incubation with a 28-amino-acid peptide representing the consensus sequence of the four GoLoco motifs of AGS3 (ref. 5). 'R12Scr' denotes pre-incubation with scrambled RGS12 GoLoco peptide⁵. **c**, $G\alpha_o$ -directed GDI activity is dependent on residues C-terminal to the 19-amino-acid GoLoco motif. Dose response curves of peptide inhibition of GTP γ S binding by Pcp2GL, R14GL, R14N/Pcp2C-GL and Pcp2N/R14C-GL are illustrated.

We previously demonstrated that RGS12, RGS14 and AGS3 GoLoco regions preferentially interact with $G\alpha_{i1}$, $G\alpha_{i2}$ and $G\alpha_{i3}$ and have no demonstrable activity towards $G\alpha_o$ ^{3,5}. The structure of the RGS14 GoLoco- $G\alpha_{i1}$ complex suggests that the sole determinants of interaction selectivity are the all-helical domain of the $G\alpha$ subunit (Fig. 1b) and GoLoco residues C-terminal to the highly conserved 19-amino-acid GoLoco motif (Fig. 3a). To test this hypothesis, we evaluated the activity of $G\alpha$ and GoLoco peptide chimaeras. RGS14- and AGS3-GoLoco peptides failed to affect GTP γ S binding to, and AIF₄⁻ activation of, a recombinant $G\alpha_{i1}$ subunit ('GiGoGi'²³) containing the all-helical domain of $G\alpha_o$ (Fig. 3b). Wild-type $G\alpha_o$ protein was also unaffected by GoLoco peptide pre-incubation, as previously shown⁵. In contrast, RGS14- and AGS3-GoLoco peptides exhibited GDI activity on a recombinant $G\alpha_o$ subunit ('GoGiGo'²³) containing the all-helical domain of $G\alpha_{i1}$, clearly indicating a role for this region of $G\alpha_{i1}$ in GoLoco binding selectivity.

We also tested the $G\alpha$ selectivity of a chimaeric GoLoco peptide representing the N terminus of the RGS14 GoLoco region and the C terminus of the Pcp2 GoLoco region ('R14N/Pcp2C-GL'; Fig. 3a). Human Pcp2 was recently shown⁶ to act as a GDI for $G\alpha_o$ (half-maximal inhibitory concentration IC₅₀ = 5.6 μ M), an activity not shared with RGS14 (ref. 5). Although somewhat less active than wild-type Pcp2 GoLoco peptide, the R14N/Pcp2C-GL peptide exhibited GDI activity towards $G\alpha_o$ (Fig. 3c), whereas a 50-fold molar excess of wild-type R14GL peptide or the reverse chimaeric peptide ('Pcp2N/R14C-GL') had no effect on GDP release by $G\alpha_o$. Given a lack of $G\alpha$ -binding specificity data for individual GoLoco motifs from the tandem-repeat proteins PINS, AGS3 and LGN, we are currently unable to discern by sequence alignments the specific residues within this C-terminal region that contribute to $G\alpha$ specificity. Additional high-resolution structures of $G\alpha_{i1}$ and $G\alpha_{i/o}$ -selective GoLoco motifs with their cognate G-protein partners will shed light on the precise structural determinants of selective $G\alpha$ interaction.

The overall structural organization of the GoLoco- $G\alpha$ complex presented here suggests that the $G\beta\gamma$ -independent GDI activity exerted on $G\alpha_{i/o}$ subunits by GoLoco proteins results from nucleotide engagement by not one, but two, arginine 'fingers'—one from each partner in the complex. On the basis of our domain-swapping experiments, we additionally propose that specific contacts between GoLoco C-terminal residues and the $G\alpha$ all-helical domain control the $G\alpha$ binding selectivity of GoLoco proteins. These results provide the mechanistic basis for further investigation to ascertain whether this GoLoco- α -GDP interaction is regulated by other signalling event(s) during the activation of signalling pathways of G-protein-coupled receptors, and/or the establishment of cell polarity, spindle organization and asymmetric cell division. □

Methods

Expression and purification

Previous structural analyses^{19,20} indicate that $G\alpha$ N termini are disordered unless bound to $G\beta\gamma$. To facilitate crystallization, we created an N-terminal-truncated, hexahistidine-tagged expression construct of human $G\alpha_{i1}$ by deleting the first 25 codons of the $G\alpha$ open reading frame within pProEX-HTb- $G\alpha_{i1}$ (ref. 5) using Quickchange site-directed mutagenesis (Stratagene). Bacterial expression, hexahistidine-tag cleavage, and purification of $G\alpha_{i1}\Delta$ N25 protein, as well as synthesis of GoLoco-motif peptides, were performed as previously described⁵.

Crystallization, structural determination and refinement

We previously delimited the RGS14 GoLoco region⁵ to a 36-amino-acid peptide ('R14GL'; amino acids 496–531 of rat RGS14) that exhibits nanomolar affinity and nucleotide-specific selectivity toward $G\alpha_{i1}$, $G\alpha_{i2}$ and $G\alpha_{i3}$ but not $G\alpha_o$. Crystals of R14GL peptide bound to α_{i1} -GDP were obtained by vapour diffusion from sitting drops containing a 1:1 (v/v) ratio of protein solution (17 mg ml⁻¹ $G\alpha_{i1}\Delta$ N25 and 1.5-fold molar excess R14GL peptide in 10 mM Tris buffer at pH 7.5, 1 mM MgCl₂, 10 μ M GDP and 5% glycerol) to well solution (100 mM sodium acetate at pH 4.6, 10% glycerol, 1.4 M caesium sulphate). Rhombohedral crystals (0.1 $\times$ 0.3 $\times$ 0.1 mm) formed in 3–5 days at 18 °C in the space group P222₁ (a = 69.5 Å, b = 82.0 Å, c = 187.2 Å, α = β = γ = 90°), with two α_{i1} -GDP-R14GL heterodimers in the asymmetric unit. For data collection at 100 K, the solution containing

the crystals was adjusted to 20% glycerol by 2% (v/v) stepwise increases in glycerol concentration. A native data set was collected on a single crystal using an R-Axis IV++ detector with Rigaku RUH3R generator and Osmic Confocal Blue Optics at the UNC-CH X-Ray Facility. All diffraction data were processed using DENZO and SCALEPACK²⁴. The structure of α_{i1} -GDP·Mg²⁺ (Protein Data Bank accession number 1BOF)¹², excluding the first 30 amino acids, water molecules and sulphates, was used as a molecular replacement model for α_{i1} -GDP-R14GL using the CCP4 program AMoRe²⁵. The program O (ref. 26) was used for model building and the program CNS (with bulk-solvent correction)²⁷ was employed for simulated annealing and torsion angle refinement. Data collection and structure refinement statistics are shown in Supplementary Information Table A. Figures 1a, c, and 2a were created using MOLSCRIPT²⁸ and Raster3D²⁹. Figure 1b was made using Spock³⁰ (<http://mackerel.tamu.edu/jon/spock>) and Raster3D.

Biochemical assays

Assays of GDI activity and GoLoco-G α binding affinity were performed as previously described². Briefly, to quantify GDI activity, initial rates of fluorescence increase on BODIPY-GTP γ S binding by G α , or activation of intrinsic G α tryptophan fluorescence by AIF₆, was measured by real-time spectrofluorometry in the presence and absence of GoLoco-containing peptide or GST-fusion protein. Binding affinity of GST-GoLoco fusion protein for α_{i1} -GDP was measured using an anti-GST surface plasmon resonance biosensor surface (BIAcore). Apparent dissociation constants were calculated using the 1:1 Langmuir (with mass transport) model as implemented by the BIAevaluation 3.0 program (BIAcore).

Received 29 October 2001; accepted 5 February 2002.

- Siderovski, D. P., Diversé-Pierluissi, M. A. & De Vries, L. The GoLoco motif: a G α -i/o binding motif and potential guanine-nucleotide exchange factor. *Trends Biochem. Sci.* **24**, 340–341 (1999).
- Takesono, A. *et al.* Receptor-independent activators of heterotrimeric G-protein signalling pathways. *J. Biol. Chem.* **274**, 33202–33205 (1999).
- De Vries, L. *et al.* Activator of G protein signalling 3 is a guanine dissociation inhibitor for G α i subunits. *Proc. Natl Acad. Sci. USA* **97**, 14364–14369 (2000).
- Natochin, M. *et al.* AGS3 inhibits GDP dissociation from G α subunits of the G_i family and rhodopsin-dependent activation of transducin. *J. Biol. Chem.* **275**, 40981–40985 (2000).
- Kimple, R. J. *et al.* RGS12 and RGS14 GoLoco motifs are G α interaction sites with guanine nucleotide dissociation inhibitor activity. *J. Biol. Chem.* **276**, 29275–29281 (2001).
- Natochin, M., Gasimov, K. G. & Artemyev, N. O. Inhibition of GDP/GTP exchange on G α subunits by proteins containing G-protein regulatory motifs. *Biochemistry* **40**, 5322–5328 (2001).
- Schaefer, M., Petronoczi, M., Dorner, D., Forte, M. & Knoblich, J. A. Heterotrimeric G proteins direct two modes of asymmetric cell division in the *Drosophila* nervous system. *Cell* **107**, 183–194 (2001).
- Brandt, D. R. & Ross, E. M. GTPase activity of the stimulatory GTP-binding regulatory protein of adenylate cyclase, Gs. Accumulation and turnover of enzyme-nucleotide intermediates. *J. Biol. Chem.* **260**, 266–272 (1985).
- Higashijima, T., Ferguson, K. M., Sternweis, P. C., Smigel, M. D. & Gilman, A. G. Effects of Mg²⁺ and the $\beta\gamma$ -subunit complex on the interactions of guanine nucleotides with G proteins. *J. Biol. Chem.* **262**, 762–766 (1987).
- Neubig, R. R. & Siderovski, D. P. Regulators of G protein signalling as new central nervous system drug targets. *Nature Rev. Drug Discovery* **1**, 187–197 (2002).
- Du, Q., Stukenberg, P. T. & Macara, I. G. A mammalian Partner of inscuteable binds NuMA and regulates mitotic spindle orientation. *Nature Cell Biol.* **3**, 1069–1075 (2001).
- Coleman, D. E. & Sprang, S. R. Crystal structures of the G protein G_{i1} complexed with GDP and Mg²⁺: a crystallographic titration experiment. *Biochemistry* **37**, 14376–14385 (1998).
- Mixon, M. B. *et al.* Tertiary and quaternary structural changes in G α induced by GTP hydrolysis. *Science* **270**, 954–960 (1995).
- Brownbridge, G. G., Lowe, P. N., Moore, K. J., Skinner, R. H. & Webb, M. R. Interaction of GTPase activating proteins (GAPs) with p21ras measured by a novel fluorescence anisotropy method. Essential role of Arg-903 of GAP in activation of GTP hydrolysis on p21ras. *J. Biol. Chem.* **268**, 10914–10919 (1993).
- Mittal, R., Ahmadian, M. R., Goody, R. S. & Wittinghofer, A. Formation of a transition-state analog of the Ras GTPase reaction by Ras-GDP, tetrafluoroaluminate, and GTPase-activating proteins. *Science* **273**, 115–117 (1996).
- Graham, D. L., Eccleston, J. F. & Lowe, P. N. The conserved arginine in rho-GTPase-activating protein is essential for efficient catalysis but not for complex formation with Rho.GDP and aluminum fluoride. *Biochemistry* **38**, 985–991 (1999).
- Peterson, Y. K. *et al.* Stabilization of the GDP-bound conformation of G α by a peptide derived from the G-protein regulatory motif of AGS3. *J. Biol. Chem.* **275**, 33193–33196 (2000).
- Paulin, F. E., Campbell, L. E., O'Brien, K., Loughlin, J. & Proud, C. G. Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. *Curr. Biol.* **11**, 55–59 (2001).
- Coleman, D. E. *et al.* Structures of active conformations of G α 1 and the mechanism of GTP hydrolysis. *Science* **265**, 1405–1412 (1994).
- Sondek, J., Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. GTPase mechanism of G proteins from the 1.7-Å crystal structure of transducin α -GDP-AIF-4. *Nature* **372**, 276–279 (1994).
- Bourne, H. R. G proteins. The arginine finger strikes again. *Nature* **389**, 673–674 (1997).
- Wall, M. A., Posner, B. A. & Sprang, S. R. Structural basis of activity and subunit recognition in G protein heterotrimers. *Structure* **6**, 1169–1183 (1998).
- Remmers, A. E., Engel, C., Liu, M. & Neubig, R. R. Interdomain interactions regulate GDP release from heterotrimeric G proteins. *Biochemistry* **38**, 13795–13800 (1999).
- Otwiniński, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunker, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Kraulis, P. J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Merritt, E. A. & Bacon, D. J. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
- Christopher, J. A. SPOCK: *The Structural Properties Observation and Calculation Kit* (Program Manual) (The Center for Macromolecular Design, Texas A&M University, College Station, 1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank R. Neubig for supplying the G α chimaeras GoGiGo and GiGoGi, members of the Siderovski and Sondek laboratories for technical assistance and support, and T. K. Harden for comments and enthusiasm. R.J.K. is supported by a predoctoral fellowship from the National Institute of Mental Health. J.S. acknowledges support from the National Institutes of Health (NIH) and the Pew Charitable Trusts. D.P.S. is a Year 2000 Scholar of the EJLB Foundation, recipient of the Burroughs-Wellcome Fund New Investigator Award in the Pharmacological Sciences, and acknowledges additional grant support from the NIH.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.P.S. (e-mail: dsiderov@med.unc.edu). The atomic coordinates have been deposited at the Research Collaboratory for Structural Bioinformatics under accession code 1KJY.

Microarray technology

An array of opportunities

technology feature

The mountain of information that is the draft sequence of the human genome may be impressive, but without interpretation that is all it remains — a mass of data. Gene function is one of the key elements researchers want to extract from the sequence, and the DNA microarray is one of the most important tools at their disposal.

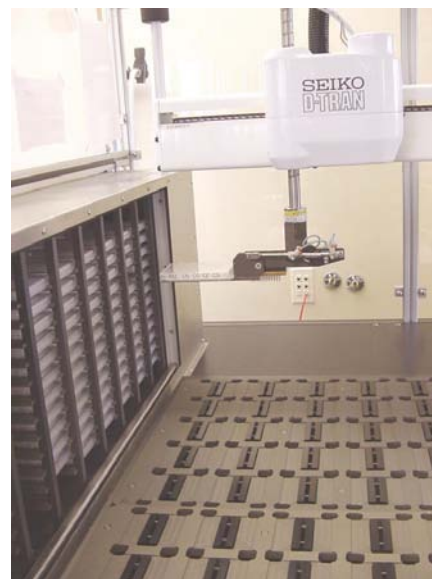
The past few years have seen rapid growth within the microarray field, with the falling price of technology allowing biologists to abandon their home-made equipment in favour of one of an expanding range of commercial instruments now on the market.

“Medicine is going molecular in a major way, and microarrays are being used to profile everything from autism and schizophrenia to Alzheimer’s and Parkinson’s,” says Mark Schena, visiting scholar at TeleChem International/arrayit.com in Sunnyvale, California, which sells microarray reagents and parts for arrayer robots.

Microarrays exploit the preferential binding of complementary single-stranded

nucleic-acid sequences. The underlying principle is the same for all microarrays, no matter how they are made — the unknown sample is hybridized to an ordered array of immobilized DNA molecules whose sequence is known. Each array features thousands of different DNA probe sequences arranged in a defined matrix on a glass or silicon support. Unlike conventional nucleic-acid hybridization methods, microarrays can identify thousands of genes simultaneously, which means that genetic analysis can be done on a huge scale.

This has revolutionized the way in which researchers analyse gene expression in cells and tissues. Microarrays — also referred to as DNA arrays, DNA chips, biochips and GeneChips — allow researchers to determine which genes are being expressed in a given cell type at a particular time and under particular conditions. They can be used to compare the gene expression in two different cell types or tissue samples, for example, healthy versus diseased tissue, and to examine changes in gene expression at different stages in the cell cycle or during embryonic development.



JEREMY HASSEMAN

Custom-built robot arrayer in action at TIGR can spot 100 slides at a time.

Microarrays are also being used in comparative genomic hybridization studies, a molecular cytogenetic approach for genome-wide detection of chromosomal deletions and amplifications, as well as for genotyping individuals for genetic differences, such as single-nucleotide polymorphisms (SNPs),

DIY OR OFF-THE-SHELF?

When microarray technology came onto the market in the late-1990s, the high price tag pushed it beyond the reach of most academic labs, so researchers were forced to use their initiative. At the time, Stanford University’s *MGuide* on how to build your own arrayer from scratch proved invaluable. Today, there are many more options.

Although it is possible to build an arrayer for about US\$50,000, a basic instrument can now be bought for about the same price from a number of suppliers — such as BioRobotics of Cambridge, UK, Genetix of New Milton, UK, Cartesian Technologies of Irvine, California, and GeneMachines of San Carlos, California — although prices for arrayers vary widely depending on their speed and capacity.

At the same time, the GeneChips made by Affymetrix are now more affordable; and ink-jet systems are starting to trickle onto the market, offering greater speed and more uniform spot morphology over contact printing, but these still come with a fairly hefty price tag.

The demand for the technology is so great at some of the major research institutions that they have established core facilities to produce inexpensive microarrays and so make the technology more broadly available. Some facilities with spare production capacity are also selling arrays at cost to investigators from outside institutions. Stanford University’s

Stanford Functional Genomics Facility, for example, offers human microarrays containing 49,000 cDNAs, of which 15,000 or more are unique human genes. The KIChip core facility at the Karolinska Institute in Stockholm offers a range of services to external researchers on a fee basis, including the production of custom-spotted microarrays.

When Vivian Cheung, of the departments of neurology and pediatric oncology at the University of Pennsylvania, Philadelphia, first thought about using microarrays in late 1996, there were no commercial arrayers and GeneChips were out of her price range. Her only option was to build a DNA arrayer in-house. Cheung’s SPOT DNA arrayer has churned out arrays for her lab for several years, where one of the main aspects of research is narrowing down the location of genes responsible for genetic diseases. She still makes and reads her own arrays, but has switched to a commercial instrument from Affymetrix. “The price is coming down to a point where it’s worth our while to buy the instruments and have someone else take care of them,” she says.

On the other hand, Michael Miles, of the department of pharmacology and toxicology at the Virginia Commonwealth University in Richmond, admits to being “sort of biased towards the commercially available arrays”, which he has been fortunate enough to be able to afford. Miles is studying the

that might be associated with disease.

At a fundamental level, microarrays are also being used in attempts to assign probable functions to newly discovered genes by comparison with the expression patterns of known genes, to identify key players in signalling pathways and to uncover new categories of genes.

But their use is not restricted to basic biology. They are also finding applications in the identification of new targets for therapeutic drugs, in disease diagnosis, and in toxicogenomics, the study of the genetic basis of an individual's response to environmental factors such as drugs and pollutants.

Spot specs

The most commonly used substrate for microarrays is glass — although they can be made of other materials, such as silicon — onto which thousands of spots of single-stranded DNA probes, in the form of cDNAs or oligonucleotides, are placed by a robot arrayer using contact or non-contact printing methods.

Alternatively, oligonucleotides can be synthesized *in situ*, building up each element of the array nucleotide by nucleotide and using ink-jet printing or photolithographic methods similar to those used in the semiconductor industry.

The spots are typically less than 200 μm in diameter and need to be read by specialized imaging equipment — confocal



Affymetrix (top) leads the high-density chip market; new kid on the block is Motorola's CodeLink chip (right).

laser scanners. The spot sizes on 'macroarrays', by contrast, are about 300 μm or more and can be imaged using conventional gel and blot scanners. Contact printing and ink-jetting methods typically give spots of 100 μm in diameter, whereas those produced by photolithography are about 20 μm . This produces microarray densities of 10,000 and 250,000 spots per cm^2 , respectively.

Industry landscape

Affymetrix of Santa Clara, California, was one of the first commercial microarray companies and still has command over the high-density microarray market. The company uses 25-mer oligonucleotides synthesized *in situ* using its proprietary process, which combines solid-phase



chemical synthesis with photolithography. Its GeneChip — an Affymetrix trademark — Human Genome U133 set of two microarrays contains over 1 million different oligonucleotides, representing more than 33,000 of the best-characterized human genes.

The price of GeneChips has come down by about half, bringing them within the reach of at least some academic researchers.

molecular plasticity of drug abuse and says DNA array studies provide a genomic-level, non-biased approach. He buys commercially produced chips but processes them on his lab's Affymetrix scanner, still a fairly expensive item at just under \$200,000.

Whether it makes sense to buy off-the-shelf or make your own arrays also depends on how many you need, and whether commercial arrays contain the genes you are interested in. But doing it yourself is not always easy. Jan Vijg, of the department of physiology at the University of Texas Health Science Center, San Antonio, says it took endless telephone calls, numerous lab visits and more than a year to develop a workable system. His research interests centre on the molecular basis of ageing and cancer. In 1999 he looked into buying commercial arrays but realized that to do large-scale

experiments of 100–200 arrays at a time would mean making his own arrays in-house. "Everything we did is really based in one way or another on information in the public domain," says Vijg. He bought a BioRobotics arrayer and an Axon Instruments scanner, and adapted the Stanford protocols, initially printing 2,000

Data analysis at the Ontario Cancer Institute's Microarray Centre.



genes per slide in duplicate. He has since been asked to turn his facility into an institutional core and has bought a second arrayer, this time from GeneMachines. This comes with a price tag of \$120,000–130,000 but has better throughput and capacity. "I think eventually we'll be able to make 20,000-gene slides in duplicate available for less than \$100," says Vijg.

When Jim Woodgett began to dabble in microarray technology three years ago, he never expected to end up running a core facility that supplies high-density microarrays and technical support to academic researchers across the globe. The Microarray Centre at the Ontario Cancer Institute in Toronto, which he directs, was established through a partnership between the institute, the government and industry to ensure that Canadian scientists had access to affordable high-quality microarrays.

Woodgett contracted with Toronto-based Engineering Services, now Virtek Vision International, to design a contact arrayer that uses a split-pin configuration. The company now sells a third-generation version of the original. Woodgett's centre runs four machines in parallel, printing from 48 genes at a time. The 30 staff generate the probes, produce the arrays and carry out quality control, and include technicians, researchers and bioinformaticians. Last year they produced 16,000 off-the-shelf arrays, including human and mouse arrays — most of which were high density, and 40% of which went to academic labs outside Ontario, many to the United States.

D.G.

The company shipped over 280,000 GeneChips in 2001 and reported revenues of US\$194.9 million, up 12% on 2000.

But recent years have seen a shake-up in the industry. Last year, Incyte Genomics of Palo Alto, California, a leading supplier of microarrays, quit the chip-making business, deciding instead to refocus its efforts on its core information business. By forging strategic collaborations with microarray manufacturers, which get access to the company's extensive database and patent portfolio, Incyte hopes to benefit from microarray sales without having to make them. Incyte may be gone, but some heavy hitters — most notably Agilent Technologies in Palo Alto and Motorola of Northbrook, Illinois — have recently entered the market-place.

It is perhaps not surprising that Motorola is making a play in this area. The company has a keen nose for business opportunities in emerging markets and the deep financial pockets needed to secure some market share. It also has core expertise in manufacturing, microfluidics, miniaturization, software engineering and systems integration.

Its subsidiary, Motorola Life Sciences, launched its first microarray product last summer. The CodeLink bioarray system for gene-expression profiling and SNP genotyping includes off-the-shelf arrays, optimized reagents and software to capture the images and carry out a first-level

analysis of the array. Labs can use their own scanners. The company offers human and rat arrays, each representing 10,000 full-length gene sequences, and expects to launch a mouse array next month. Its genotyping array contains 72 SNPs from the P450 cytochrome family. Motorola's agreement with Incyte Genomics allows it to develop microarrays based on Incyte's comprehensive gene databases.

Motorola synthesizes 30-mer oligonucleotides 'off-line' and spots them onto slides coated with a three-dimensional, branched polymeric substrate gel surface, using Hewlett-Packard's non-contact, piezo-dispense technology. The company also produces custom arrays to order and sells 'activated' non-spotted slides for researchers to make their own arrays.

Agilent Technologies, on the other hand, uses proprietary SurePrint ink-jet technology and offers human, mouse and rat cDNA arrays and custom oligonucleotide arrays. In the latter case the oligonucleotides (either 25- or 60-mer) are synthesized *in situ* and built up a base at a time on standard 1 × 3-inch glass slides to give arrays of either 8,400 or 22,000 features. Doug Amorese, R&D section manager responsible for chemistry and molecular biology in Agilent's DNA Microarray Program, says the cDNA type of microarray is useful when large numbers of identical arrays are needed,

whereas the *in situ* system provides the flexibility to tailor designs to suit individual needs.

As a subsidiary of Hewlett-Packard, Agilent has access to considerable expertise in ink-jet printing methods and high-end analytical instrumentation — principally high-performance liquid chromatography and mass spectrometry. So the microarray area "seemed like a very good fit" for the company, says Amorese. Hewlett-Packard had been looking for a way into molecular biology, and microarrays "seemed like an area that was going to grow", he says.

The cross-licensing agreement Agilent signed in 1999 with Oxford Gene Technology (OGT) of Oxford, UK, is seen by the company as key to making this happen. OGT was set up by Edwin Southern and the University of Oxford in 1995 to commercialize Southern's DNA microarray patents. Agilent's other main collaborators are Rosetta Inpharmatics of Kirkland, Washington, and Incyte Genomics.

David and Goliath

As well as the big guns, several smaller companies are seeking to carve out a niche. One example is febit, a young biotechnology company employing some 70 people in Mannheim, Germany. It has developed a prototype DNA analysis device that fully automates and integrates all the steps in the analysis process. Its machine, Geniom one, is designed for both gene-expression analysis

DEALING WITH THE DATA DELUGE

The massive amount of microarray data collected so far has been generated on multiple platforms and is stored in a host of different formats, levels of detail and locations. This makes it difficult for any group to re-analyse or verify the data, or compare the results with their own. "It's apples to oranges," says Steven Gullans of the department of medicine at Brigham and Women's Hospital/Harvard Institutes of Medicine in Boston, Massachusetts.

Moreover, there are no uniform standards for reporting microarray data in journal articles, and there is no requirement for authors to deposit their data — and any supporting information — in the public domain. "I think the journals have to force it," says Gullans, "just like they forced us to put sequence data in the public databases, and they are a little at a loss how to do that."

Although most researchers agree that public databases for microarray data are a good idea, many are hesitant about depositing their own data in the public repositories now being developed. These include the Gene Expression Omnibus (GEO), operated by the US National Center for Biotechnology Information (NCBI); ArrayExpress, run by the European Bioinformatics Institute (EBI) in the UK; and CIBEX, the gene-expression database being developed by the DNA Data Bank of Japan.

"I think everyone realizes that the value of [microarray] data is not in looking at them in isolation but really trying to look at them in a broader context," says John Quackenbush, head of the whole-

genome functional analysis group at The Institute for Genomic Research in Rockville, Maryland.

The problem is that expression data are much richer than sequence data, and many factors can affect how genes are expressed. You need to capture more information, says Quackenbush, including details of the experimental design, array design, samples, controls and experimental conditions, and the data manipulation and analysis methods used.

The Microarray Gene Expression Data (MGED) group was established in 1999 to develop a framework for describing information about a DNA microarray experiment, as well as a standard format for data exchange. The first version of its MIAME (minimum information about a microarray experiment) was proposed last year (see *Nature Genet.* 29, 365–371; 2001 and *Nature* 415, 946; 2002). The MAGE-ML (Microarray Gene Expression Markup Language) data-exchange format, which the MGED is developing along with the

Data, data everywhere



JEREMY HASSEMAN

and genotyping. It offers “a plug-and-play solution”, says Peer Stähler, febit’s vice-president and chief scientific officer, and one of the company’s founders. “You don’t have to become an expert in surface chemistry, you don’t have to optimize the processes. All you need is data,” he says.

At the heart of Geniom one is the programmable DNA processor — a special reaction carrier with a three-dimensional microchannel structure. Both the synthesis of the oligonucleotide probes — which uses a light-dependent technique that does not rely on physical masks — and the hybridization of the labelled samples takes place in the channels. “You insert the reaction carrier and never touch it again until you throw it away,” says Stähler. “If you’re efficient you can do two runs a day.”

The current design can produce microarrays containing up to 64,000 different oligonucleotides — it runs eight arrays in parallel, each with 8,000 spots per array. With between one and four spots covering a gene, each array can cover a few thousand genes. This is not as dense a coverage as Affymetrix’s GeneChips, but Stähler expects future versions of Geniom to have 10 times as many spots per array.

The prototype is being tested by Jörg Hoheisel and his team at the German Cancer Research Centre in Heidelberg. Stähler expects Geniom one, which has a price tag of a few hundred

thousand dollars, to hit the market by the end of the year.

Room for improvement

There is still a lot of room for improvement in microarray technology, say players in the field. TeleChem International/arrayit.com, for example, is exploring the use of reflective substrates. Although still in the development phase, Schena says it seems that printing microarrays on mirrors rather than glass improves the signal-to-noise ratio by as much as 1,000%.

Several companies are pursuing the development of ‘active’ hybridization technologies. Advantix, a recent spin-off from the Center for NanoScience at the Ludwig-Maximilians University of Munich, will begin shipping a hybridization device this month, which has no moving parts and is designed to speed up hybridization reactions, as well as to produce more homogeneous reaction conditions than with ‘passive’ hybridization, eliminating so-called edge effects. The mixer chip uses surface acoustic waves to control the motion of reagents. It is used in a ‘sandwich’ arrangement, with a conventional DNA microarray slide on the bottom, the mixer chip on top and the hybridization solution in between.

“Microarrays will get better over time and a lot of that will be in content as we



Peer Stähler (left) and board members of febit with Geniom one.

FEBIT

better understand which genes are important and, specifically, perhaps which splice variants are most important,” says Amorese. In addition to improvements in the probes themselves, he expects advances in labelling technologies for the sample nucleic acid, allowing researchers to use less starting material. As for chips in the clinic, Schena believes they will be there within five years, and probably a lot sooner on the genetic screening side.

Diane Gershon is Assistant Editor, New Technology at *Nature Medicine*.

- ◆ www.microarray.org
- ◆ www.gene-chips.com
- ◆ www.lab-on-a-chip.com
- ◆ cmgm.stanford.edu/pbrown/mguide/index.html
- ◆ www.mged.org/

Life Sciences Research Task Force of the Object Management Group (OMG), a software standards organization, moved a step closer to implementation after a recent vote within the OMG.

“It all boils down to whether we want to continue in the life sciences with a tradition that the supporting data should be available, or not,” says Alvis Brazma, team leader for microarray informatics at the EBI. Brazma is responsible for spearheading efforts to adopt minimum standards for microarray data and a standard data-exchange format.

The MGED has sought the input of the microarray community, including software and hardware companies. Rosetta Inpharmatics, for example, was working on its own standard, but has since joined forces with the MGED. “Our goal was to have a standard that everyone would use and that was at risk if we had a lot of smart folks working on two different applications,” says Doug Bassett, vice president and general manager of Rosetta Biosoftware, the recently formed software arm of the company. Bassett expects the company’s software products, which include the Rosetta Resolver gene-expression data analysis system, to be among the first to offer full support for MAGE-ML.

EBI’s ArrayExpress currently houses only three data sets, but it now accepts data in the MAGE-ML format. The EBI is beta-testing the web-based data submission capabilities for ArrayExpress, and Brazma expects this phase to last another 2–3 months.

The GEO, launched by the NCBI last July, has been operational for longer, contains more data, and both accepts data submissions

JEREMY HASEMAN



Quackenbush: supports data standards

and supports data queries. But some researchers find it difficult to work with. “GEO has the disadvantage that all of the data are stored basically as a big tab-delimited file inside the database. That makes it very difficult to query,” says Quackenbush. The NCBI is developing a set of tools on top of the GEO to try to extract the information and make it more accessible. Yoshio Tatenos, of the Center for Information Biology, part of the National Institute of Genetics in Mishima, Japan, expects CIBEX to be publicly accessible and support MAGE-ML some time this summer.

Some private databases are also working towards supporting MAGE-ML and being MIAME-compliant. Gavin Sherlock, director of Microarray Informatics at the Stanford Microarray Database, hopes the database will be MIAME-compliant by the end of this year. “One of the things that makes it hard for us is the quantity of data we already have,” he says, which amounts to information from some 22,000 arrays.

The MGED is also about to come up with a checklist for authors, editors and reviewers of what information should be given in microarray-based papers and what supporting information should be revealed electronically — details of which will be posted on its website. Brazma hopes it will serve as a useful guide that “will put everything on a more level playing field”. D.G.

Company	Products/activity	Location	URL
General: microarray systems, arrays and oligo synthesis			
Advanced Array Technology	Low-density microarrays	Namur, Belgium	www.aat-array.com
Affymetrix	GeneChip arrays and reagents, instrument systems and software	Santa Clara, California	www.affymetrix.com
Agilent Technologies	Catalogue and custom arrays, scanner, software	Palo Alto, California	www.agilent.com
Alpha Technology	Customized low- and medium-density microarrays	Hürth, Germany	www.alpha-tec.net
Amersham Biosciences	Lucidea array spotter, automated slide processor, scanner, spotfinder, scoring system	Uppsala, Sweden	www.apbiotech.com
BD Biosciences Clontech	Atlas glass/plastic/nylon arrays, labelling kits, AtlasImage/AtlasNavigator software	Palo Alto, California	www.clontech.com
BioCat	Distributor of catalogue and customized arrays	Heidelberg, Germany	www.biocat.de
Exiqon	Microarray slides and chips	Vedbaek, Denmark	www.exiqon.com
febit	Geniom one system for gene-expression profiling or genotyping	Mannheim, Germany	www.febit.com
GeneScan Europe	Biochip technology; high-throughput production of biochips	Freiburg, Germany	www.genescan-europe.com
Genisphere	3DNA Submicro Expression Array labelling and detection kits; microarray services	Hatfield, Pennsylvania	www.genisphere.com
Genomic Solutions	GeneTAC biochip system/GeneMAP pre-printed arrays	Ann Arbor, Michigan	www.genomicsolutions.com
Greiner Bio-one	Ready to use biochip kits for genotyping and SNP detection, modified slides for microarraying	Frickenhausen, Germany	www.greinerbioone.com
Illumina	Oligator custom DNA synthesis services/SNP genotyping services	San Diego, California	www.illumina.com
Lambda	Ready to use biochip kits for genotyping and SNP detection	Freistadt, Austria	www.lambda.at
Motorola Life Sciences	CodeLink bioarray system: arrays for gene-expression profiling/SNP detection, software, activated slides	Northbrook, Illinois	www.motorola.com/lifesciences
MWG Biotech	Oligo-based microarrays, custom chips, automated systems	Ebersberg, Germany	mwgatccn.mwgdna.com/
Nanogen	NanoChip automated workstation for SNP and STR analysis	San Diego, California	www.nanogen.com
Orchid BioSciences	Array-based SNP genotyping platforms: instruments and consumable kits for SNP scoring	Princeton, New Jersey	www.orchid.com
Oxford Gene Technologies	Development of customized high-density microarray technology and services	Oxford, UK	www.ogt.co.uk
Packard BioScience (Division of PerkinElmer Life Sciences)	Hydrogel-coated slides, arrayers, scanner and data-analysis software	Meriden, Connecticut	www.packardbioscience.com
PamGene	Low-density microarray systems	Hertogenosch, The Netherlands	www.pamgene.com
PanVera	Intelligene microarrays	Madison, Wisconsin	www.panvera.com
PerkinElmer Life Sciences	Micromax high-throughput human DNA microarrays	Boston, Massachusetts	lifesciences.perkinelmer.com
PicoRapid Technologie	PicoArrays, PicoSlides and microarray spotting service	Bremen, Germany	www.picorapid.de
Qiagen Operon	Custom arrays and Array-Ready oligo sets	Alameda, California	www.operon.com
ResGen Invitrogen	GeneFilters microarrays/VastArray tissue arrays/ GeneStorm expression-ready full-length clones	Huntsville, Alabama	www.resgen.com
Schleicher & Schuell	PIQOR technology for producing custom and generic murine and human cDNA arrays	Dassel, Germany	www.s-und-s.de
Scienion	Generic and customized microarrays	Berlin, Germany	www.scienion.de
Sigma-Genosys	Custom oligos, oligo libraries, DNA arrays	The Woodlands, Texas	www.sigma-genosys.com
SmartBead Technologies	3D UltraPlex™ microarray technology for applications in SNP genotyping and gene expression	Cambridge, UK	www.smartbead.com
SuperArray	Gene-expression array, GEArray systems designed for pathway-specific gene expression profiling	Bethesda, Maryland	www.superarray.com
VBC-GENOMICS	Custom oligonucleotide microarrays and Affymetrix GeneChip services	Vienna, Austria	www.vbc-genomics.com
Vysis	Microarrays, microarray readers and reagents	Downers Grove, Illinois	www.vysis.com
Robotics			
BioRobotics	Arrayers	Cambridge, UK	www.biorobotics.com
Cartesian Technologies	Contact and non-contact arrayers, Clone Tracker software	Irvine, California	www.cartesiantech.com
GeneMachines	OmniGrid and OmniGrid Accent microarrayers	San Carlos, California	www.genemachines.com
Genetix	Arrayers, scanners and consumables	New Milton, UK	www.genetix.co.uk
MiraBio (Hitachi Genetic Systems)	Microarray systems, bead-based liquid array system, SPBIO II spotter and CRBIO lie scanner, software and consumables	Alameda, California	www.miraibio.com
QIAGEN	Custom oligos, SNP analysis, molecular biology workstations (BioRobot 3000/BioRobot 8000)	Valencia, California	www.qiagen.com
Radius Biosciences	3XVP arrayer, custom microarrays	Medfield, Massachusetts	www.ultranet.com/~radius
RELAB	Arrays, scanners, workstations and dispensers for biochips	Recklinghausen, Germany	www.relab.de
RoboDesign International	RoboArrayer industrial microarrayer with built-in QC	Carlsbad, California	www.robodesign.com
Tecan	Robotic laboratory workstation	Männedorf, Switzerland	www.tecan.com
TeleChem International/arrayit.com	Fine chemicals and manufacturer of ArrayIt products for microarray analysis	Sunnyvale, California	www.arrayit.com
Virtek Vision International	ChipWriter Pro high-throughput DNA-spotting system	Waterloo, Canada	www.virtekvision.com

Scanners and imaging systems

Alpha Innotech	Microarray scanners	San Leandro, California	www.alphainnotech.com
Applied Precision	arrayWoRxe biochip reader	Issaquah, Washington	www.appliedprecision.com
Axon Instruments	GenePix 4000B microarray scanner and GenePix Pro 4.0 microarray analysis software	Union City, California	www.axon.com
Kodak	Imaging systems and image analysis software	Rochester, New York	www.kodak.com
LaVision Biotec	BioAnalyzer biochip readers	Bielefeld, Germany	www.lavisionbiotec.de
Polaroid	Scanners and imaging systems	Cambridge, Massachusetts	www.polaroid.com
UVP	Bioimaging systems	Upland, California	www.uvp.com
Nikon	Scanners and imaging systems	Tokyo, Japan	www.nikon.com
Olympus	Scanners and imaging systems	Tokyo, Japan	www.olympus.com

LCM-microarrays

Arcturus	Microgenomics system using laser-capture microdissection	Mountain View, California	www.arctur.com
----------	--	---------------------------	--

Software

Biocomputing	Software and bioinformatics services for genotyping	Espoo, Finland	www.biocomputing.fi
BioDiscovery	Software products for microarray research and analysis (CloneTracker, ImaGene, GeneSight, GeneDirector)	Marina del Rey, California	www.biodiscovery.com
GeneData	Bioinformatics systems and services for functional genomics	Basel, Switzerland	www.genedata.com
Imaging Research	Microarray analysis and statistical validation software	Ontario, Canada	imaging.brocku.ca
InformMax	Xpression NTI/Array-Pro Analyzer/GenoMax and BioRS	Bethesda, Maryland	www.informaxinc.com
IBM Life Sciences	Data management systems	White Plains, New York	www.ibm.com/solutions/lifesciences
Jobion Informatics	GeneTraffic microarray data management and analysis software	La Jolla, California	www.iobion.com
LION Bioscience	Software solutions for entire R&D chain	Heidelberg, Germany	www.lionbioscience.com
Molecular Mining	Data mining software products	Kingston, Ontario, Canada	www.molecularmining.com
Rosetta Biosoftware	Rosetta Resolver gene-expression data analysis system	Kirkland, Washington	www.rosettatabio.com
Scanalytics	MicroArray Suite	Fairfax, Virginia	www.scanalytics.com
Silicon Genetics	MetaMine, GeNet and GeneSpring microarray software	Redwood City, California	www.sigenetics.com
Spotfire	DecisionSite software	Somerville, Massachusetts	www.spotfire.com
X-MINE	XOL, Opus, X-Miner, X-Reporter software	Brisbane, California	www.xmine.com

Databases

Gene Logic	Gene-expression databases derived from clinically important tissues	Gaithersburg, Maryland	www.genelogic.com
Incyte Genomics	Databases	Palo Alto, California	www.incyte.com

Kits

Ambion	RNA-based life-science research and molecular biology products	Austin, Texas	www.ambion.com
Millipore	96-well purification plates for PCR; PCR clean up	Bedford, Massachusetts	www.millipore.com
RNAture	mRNA isolation/cDNA synthesis kits	Irvine, California	www.rnature.com

Slides

Coming	GAPS II microarray slides, hybridization chambers	Coming, New York	www.coming.com
Eppendorf	Coated slides for arrays; molecular biology reagents	Hamburg, Germany	www.eppendorf.com
Memorec Stoffel	Coated slides for microarrays; nylon membranes for PCR-based arrays	Cologne, Germany	www.memorec.com
Mosaic Technologies	EZ-RAYS Activated slide kits for DNA microarrays	Waltham, Massachusetts	www.mostek.com
Scandinavian Micro Biodevices	Slides for DNA arrays	Lyngby, Denmark	www.smb.dk
SurModics	3D-Link activated slides	Eden Prairie, Minnesota	www.surmodics.com

Services

Expression Analysis	GeneChip processing and data analysis	Durham, North Carolina	www.expressionanalysis.com
---------------------	---------------------------------------	------------------------	--

Miscellaneous

Advantix	Microfluidic processors	Munich, Germany	advantix.de
Berthold Technologies	Bioanalytical instrumentation	Bad Wildbad, Germany	www.bertholdtech.com
Graffinity Pharmaceuticals	Preclinical drug discovery	Heidelberg, Germany	www.graffinity.com
Molecular Machines & Industries	Bioanalytical consumables, single-molecule counters and assay automation facilities	Heidelberg, Germany	www.molecularmachines.com
SuNyx Surface Nanotechnologies	Microfluidic devices/biochips	Cologne, Germany	www.sunyx.de
Thermo LabSystems	Laboratory information management systems	Altrincham, Cheshire	www.thermolabsystems.com

● See advertisement

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Laura Pearson (4977)
Netherlands/ Italy/ Iberia:
Evelina Rubio Hakansson (4973)
Scandinavia: Silke Opstrup (4994)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:
Hind Berrada (4935)

Naturejobs web development:
Tom Hancock

Naturejobs online production:
Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakata-machi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com
Japan Manager: Kate Cowan

naturejobs

Going into industry?

Although academic life scientists are aware of industrial opportunities, they are also wary of them. So says a recent poll on the perceptions of the advantages and disadvantages of working in the pharmaceutical industry, conducted by the Science Advisory Board, an online forum of researchers in the life and medical sciences.

Of the 778 respondents, about a third cited an inflexible corporate culture or the inability to set their own research priorities as the main drawbacks of a career in the pharmaceutical industry. Another 19% mentioned restrictions on publication as the main disadvantage.

But the respondents were also aware of the advantages of industry. They noted that companies offer a team-oriented environment with multiple career paths. Also mentioned was the fact that companies often have better research facilities, with equipment that can increase research productivity.

These pros and cons are echoed by postdocs working in industrial laboratories (see page 5). But their stories also show that policies on research autonomy and publication vary widely from company to company.

All this adds up to a lesson in preparation. Anyone — postdoc or otherwise — considering industrial employment would be well advised to investigate their potential employer's publication record, talk to former and current employees and ask pointed questions about research autonomy. Then, a decision to join a company can be made with confidence rather than caution.

Paul Smaglik
Naturejobs editor



Contents

POSTDOCS & STUDENTSHIPS

Choosing a good company p5

MOVERS

English cancer researcher leaves
NIH to head Scottish lab;
British astronomer leaves
Canada for Switzerland; and
more [Back page](#)

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Getting into good company

Mariano Troccoli:

Being in a company, in a hierarchical structure, gives you more motivation to achieve your goals.



LUCENT

New PhDs used to approach industrial postdoctorates with some trepidation. Such positions had a reputation for allowing little, if any, publication. PhDs who accepted them felt that they were sealing off their options of any future academic career. But although these issues warrant consideration, they are not true in every case.

Publication records vary from company to company. And for people willing to do multiple postdocs, an industrial fellowship before or after an academic one offers a way to explore the idea of a career in industry without committing to it.

PUBLICATION CONSIDERATION

Prospective industrial postdocs may want to investigate a company's policy and see what other postdocs have published while in similar positions. Some companies have a history of their scientists publishing, such as IBM's nanotechnology research section in Yorktown Heights, New York, and in Zurich, Bell Labs in Murray Hill, New Jersey, and Genentech in South San Francisco.

That record played a factor in Sachdev Sidhu's decision to accept a postdoc with Genentech — a decision that proved sound. "I published more than 10 papers in the past two years, more than I could have published in an academic setting," Sidhu says.

Unlike those in academia, an industrial postdoc can lead directly on to a job in the same place. Both Sidhu and his colleague Sarah Hymowitz accepted offers from the company after completing fellowships there. Biotech firms are more likely to hire their own former postdocs than are drug companies, but policies vary and exceptions are often made for valuable postdocs.

Another question to ask is how long-term and basic a research project will be. Short-term, applied goals in industrial research can be problematic for people with a more academic bent. Kurt Lust, now a researcher at the University of Leuven in Belgium, did a postdoc at United Technologies in Hartford, Connecticut, conducting numerical bifurcation analysis. "We were busy with some nice things, but never had the time to demonstrate that they were nice," says Lust.

Companies that include long-term and even basic science projects can offer more interesting postdoc

programmes. Martin Bammerlin, who is involved in nanoscience research at IBM's Research Laboratory in Zurich, says that he is doing basic science, not projects involving building countless prototypes.

Long-term research often depends on how much money a company can invest in research. "At Bell Labs, basic research is still an important issue. This is something that is essential for the company," says Mariano Troccoli, who has a two-year contract in Bell Labs' semiconductor physics research department.

Many industrial postdocs pick up skills or experiences they would not have received in academia. Bernd Gotsmann, who does nanoscience research as a postdoc at the IBM Zurich Research Laboratory in Switzerland, says that some of the experience he has picked up there may be valuable in academia — especially being exposed to teamwork. He has also enjoyed interacting with people from different fields.

DIFFERENT APPROACHES

Gotsmann, who is a physicist, worked with people from the computer sciences, engineering and chemistry. "They all have different approaches to the problems," says Gotsmann. Those interactions could be at a premium if he returns to academia: many projects are becoming increasingly interdisciplinary, even though few academics have the experience of interacting outside their departments.

Project and time management skills also tend to be emphasized more in industry. This is especially true for big projects, says Marina Brockway, who did a PhD at Kiev University and a postdoc at the University of California, Irvine, before joining Honeywell of Minneapolis, Minnesota, in an industrial postdoc programme. "Trying to resolve a problem in a timeline, trying to make a problem manageable" is often lacking in academic research, she says.

That environment may not be for everyone. But being exposed to it — whether or not you intend to stay in it — can be valuable, says Troccoli, who is enjoying the fast-paced environment of his Bell Labs fellowship. "Being in a hierarchical structure in a company gives you more motivation to achieve your goals," he says.

Alexander Hellemans is a science writer in Naples, Italy.

Web links

Bell Labs
 ♦ www.bell-labs.com
 IBM Zurich
 ♦ www.zurich.ibm.com
 IBM at Yorktown Heights
 ♦ www.watson.ibm.com
 Genentech
 ♦ www.gene.com/gene/research
 United Technologies
 ♦ www.utrc.utc.com
 Honeywell
 ♦ www.honeywell.com/index.html

Transitions

X Merrimack Pharmaceuticals, a Cambridge, Massachusetts biotech company, this month appointed **Mark Perrone** as vice-president of research and development. Perrone comes from Bristol-Myers Squibb.

X Cozart Bioscience, an Oxfordshire-based medical diagnostics company, last month appointed **David McGough** as international business development manager. McGough comes to the company from LifePoint.

X Princeton-based Orchid BioSciences this month appointed **Kenneth Noonan** to its board of directors. Noonan is a partner at London's LEK Consulting, where he leads the firm's life sciences practices in the United Kingdom and Europe.

X Paris-based functional-proteomics company Hybrigenics this month hired **Anne-Fabienne Weitsch** as chief operating officer. Weitsch was previously senior director of licensing and business development at Johnson & Johnson.

X **Mike Pavia**, chief technology officer for Millennium Pharmaceuticals, has joined the board of microfluidics company BioProcessors of Woburn, Massachusetts.

X This summer **Peter Baumann**, a postdoc at the University of Colorado, Boulder, will join the Kansas City-based Stowers Institute, a basic cell biology institute which opened last year.

X This month Astex Technology, a UK protein-structure-based drug-discovery company, appointed **Peter Fellner** as non-executive chairman of the board. Fellner has been CEO of Celltech Group since 1990.

CANCER RESEARCH

Karen Vousden, chief of the Regulation of Cell Growth Laboratory at the US National Cancer Institute (NCI), will return to the United Kingdom this summer to help expand a cancer research centre in Scotland. Vousden was appointed director of the Beatson Institute in Glasgow last month. Part of the attraction is a £10 million (US\$14.3 million) expansion of the Beatson's campus in Bearsden, which will provide facilities for 100 additional scientists when it is completed in 2004. The construction will be a big draw to prospective scientists, says Vousden, who is looking forward to recruiting. Some postdocs from her NCI lab have already agreed to make the move with her in August. "Others are thinking about whether they can bear the rain," she jokes.

PHYSICS

Associate professor of physics **Dik Bouwmeester** this month took up permanent residence in his new lab at the University of California, Santa Barbara, after moving from Oxford in January.

Bouwmeester, who studies the quantum properties of light and is investigating how those properties could be used in cryptography, was attracted to California in part because the university system there is committed to increasing its investments in nanotechnology — including \$200 million being distributed between the Santa Barbara and Los Angeles campuses for new facilities.

He hopes that those facilities, and interactions with a growing pool of scientists who will fill them, will provide the opportunity to combine his interests in optics and nanotechnology.

ASTRONOMY

After a little over a year as director-general of Canada's centre of astronomy, the Herzberg Institute of Astrophysics (HIA) in British Columbia, **Simon Lilly** has left. Lilly departed in February to take up a position as professor at the Institute for Astronomy of the Swiss Federal Institute of Technology (ETH) in Zurich.

Many factors, "both personal and professional" contributed to his decision. On the personal level, the London-born astrophysicist had been in Canada since 1990. "The thought of coming back to Europe was really attractive," Lilly says.

On the professional level, he felt he accomplished his main goal at the HIA, securing the future of the Long Range Plan for Canadian astronomy, developed in 1998. "I felt I had already set things on a very positive track," Lilly says. Although he



Karen Vousden



Simon Lilly



Paul Wellings



Brian Popko

enjoyed his administrative duties with the HIA, he says he is keen to focus more on science, particularly in increasing the ETH's presence in extragalactic astrophysics and observational cosmology.

ENVIRONMENTAL SCIENCE

After 21 years with Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO), British environmental scientist **Paul Wellings**, the organization's deputy chief executive, is leaving Canberra to return to Britain. In August he takes up the position of vice-chancellor at Lancaster University.

Wellings previously held positions as chief of CSIRO entomology and head of the science and innovation division in the Australian Department of Industry, Science and Resources before becoming a member of the CSIRO executive in 1999. Wellings will succeed **Bill Ritchie**, who retires this summer.

NEUROBIOLOGY

This month, neurobiologist **Brian Popko** was appointed the first Jack Miller professor in peripheral neuropathy at the University of Chicago. Popko researches myelinating glial cells, which provide a protective 'sheath' around nerves, and interactions between the immune and nervous systems.

Popko came to Chicago from the University of North Carolina at Chapel Hill, where he was a professor of biochemistry and biophysics and director of the functional genomes core facility.

ADMINISTRATION

Russel Kaufman has been appointed director and chief executive of the Wistar Institute, an independent non-profit biomedical research centre in Philadelphia. Currently a vice-dean at Duke University, Kaufman has been at the university since his medical residency in 1973, with the exception of a two-year break to do a fellowship at the US National Institutes of Health.

The new job will allow him to focus his interests back on basic biomedical research. At Duke, he is responsible for educational and academic affairs. "Right now I have a really broad palette," he says. At Wistar, he wants to differentiate the institute from the University of Pennsylvania campus nearby, while at the same time strengthening the ties between the two institutions.

Several faculty at Wistar have joint appointments with the University of Pennsylvania, but Kaufman would like to see more formal research collaborations forged between the two.

CONTACTING US AT MOVERS

Please send any information on appointments or job moves to:
▶ naturejobseditor
@naturedc.com.